The PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

L. 43 SEVENTH SERIES No. 346 November, 1952

A Journal of Theoretical Experimental and Applied Physics

PRICE 15s. 0d.
Annual Subscription £8 0s. 0d. payable in advance

AND PUBLISHED BY TAYLOR & FRANCIS LTD., RED LION COURT, FLEET ST., LONDON, E.C.4.

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT OF THE PHILOSOPHICAL MAGAZINE

On 1st January, 1952, the first number of this new Quarterly Supplement to the Philosophical Magazine was published. The aim of this Supplement will be to give those interested in physics comprehensive and authoritative accounts of recent important developments. It is felt by the Editor that in view of the rapid advances in many branches of physics, scientists will welcome a journal devoted to articles of this type.

CONTENTS FOR VOLUME 1, 1952 NOS. 1-4

PRICE *per part* **15/-** *plus postage* **PRICE** *per annum* **£2 15s. 0d.** *post free*

Printed and Published by
TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET ST., LONDON, E.C.4

CXIV. *X-ray Transition Probabilities with Special Reference to K Absorption in Lithium*

By J. Friedel
H. H. Wills Physical Laboratory, University of Bristol*

[Received August 8, 1952]

Summary

To interpret the position of x-ray absorption edges in monovalent metals, the author assumed in a previous paper that the positive hole created in an inner shell was screened by a bound electron. It is shown here that this assumption gives also a correct order for the magnitude of the absorption coefficient near the edge.

In an introductory treatment we take the absorbing system to be the ejected electron alone. We use the hydrogenic wave-function for the bound state and the plane wave for the continuum. We thus obtain approximate formulae for the total jump in the absorption coefficients at the K, L and M edges, both for x-rays and in the optical region.

In a further treatment, taking lithium as an example, we treat the absorbing atom and the Fermi electrons as a whole, and with this model analyse the fine structure of the edge. We show how the screening may be obtained from the conduction band. We analyse various possible types of transitions producing edges or lines in the absorption, and relate them to the structure observed near the edge. To avoid too much complexity in treating the Fermi electrons, we enclose the metal in a large but finite sphere. Quantitative agreement with experiment is not altogether satisfactory, possibly because of inaccuracies in the wave-functions used.

Finally, we describe briefly the reverse process of K-emission in lithium, and discuss the possible presence of lines as well as bands in the x-ray spectra of monovalent metals.

§1. Introduction

A theoretical treatment of photoelectric absorption by free atoms in the non-relativistic region has been given by Stobbe (1930). Stobbe's very simplified model consists of one electron in the field of a bare nucleus of charge Z. The wave-functions of both the initial bound state and the final free state are then hydrogenic ; and an exact computation gives for the K-absorption coefficient (in cm^{-1}):

$$K=8\pi N\phi_0(137)^3\frac{Z^6}{(h\nu)^4}\frac{\exp(-4\,\xi\,\mathrm{arccot}\,\xi)}{1-\exp(-2\pi\xi)}. \quad . \quad . \quad . \quad (1)$$

* Communicated by Professor N. F. Mott, F.R.S.

Here N is the number of atoms per cm^3; $\phi_0=8\pi e^4/3m^2c^2$ is the classical Thomson scattering factor; the energies $h\nu$ and E are expressed in atomic units*; and $h\nu=\frac{1}{2}Z^2+E$ is the energy of the photon, where E is the kinetic energy of the expelled electron; finally $\xi^2=\frac{1}{2}Z^2E^{-1}$.

To apply formula (1) and a similar one for $L_{II,\,III}$ to actual atoms, Stobbe makes two corrections. He replaces Z by an 'effective' charge Z_e which takes into account the screening of the charge acting on the K-electron by the other electrons ($Z_e=Z-0{\cdot}3$). He also uses for $h\nu$ the experimental energy of absorption instead of $\frac{1}{2}Z^2+E$. Good agreement with experiment is then obtained for *solid* absorbers (Stobbe, *loc. cit.*; cf. also Bethe 1933 a, and Heitler 1936 a). Some of Stobbe's examples show, however, large deviations near the edges, and agreement is especially poor for light elements. For (solid) lithium, for instance, $h\nu=54{\cdot}5$ ev (Skinner and Johnston 1937) and at the edge $E\simeq 4{\cdot}7_5$ ev (Mott and Jones 1936). With these values formula (1) predicts a jump of $86{\cdot}5\times10^4$ cm^{-1}; the experimental measurement (Skinner and Johnston, *loc. cit.*†) gives about 3×10^4. The theoretical value is not very sensitive to E.

It is therefore of interest to analyse the absorption process more closely. Two criticisms can be made of Stobbe's treatment: the field acting on the expelled atom is certainly less attractive than Z_er^{-1}; also the wave functions of the remaining electrons are altered by the absorption process. These two effects are most important for light elements near the absorption edge. We shall study them in turn, restricting ourselves in the second case to the simple case of lithium K absorption.

§ 2. Screening of the Positive Hole

An improvement on Stobbe's treatment for a gas would be to include the effect of all the other electrons by treating the absorbing electron as moving in a field of which the potential behaves like Zr^{-1} near the nucleus and r^{-1} at large distances. In a metal, however, the effect of all the other electrons is more complicated. The conduction electrons after a K-absorption process must be in equilibrium in the field of the positive hole which replaces the missing K electron. The motion through the lattice of this hole is slow and can be neglected; and for transitions near the absorption edge, its coulomb field must be *screened* by the mobile conduction electrons. The screening, after absorption at the edge, is of the type most favourable energetically. In lithium, this is provided by a 2s bound electron; and the energy of the edge agrees satisfactorily with the energy of excitation of the Li^+ ion from the $1s^2$ to the 1s2s state (Friedel 1952 a).

In a metal, therefore, near the absorption edge, it seems a better approximation than Stobbe's is to assume that the expelled electron

* Atomic units are used throughout this paper.

† There is an obvious error of sign in the value reported on the figure.

moves in the unperturbed lattice. Neglecting the lattice structure, we therefore write for its wave function a plane wave exp ($i\mathbf{k}\,.\,\mathbf{r}$), and for the 1s electron a hydrogenic wave function :

$$1s=\pi^{-1/2}Z_e^{3/2}\exp(-Z_e r). \qquad (2)$$

Let $Op=\exp(-i\boldsymbol{\varkappa}\,.\,\mathbf{r})k_e$ be the operator of the transition (Mott and Sneddon 1948). In this expression, k_e is the projection along the direction of polarization $\mathbf{e}$ of the momentum $\mathbf{k}=-i\nabla$ of the expelled electron, considered as an operator. $\boldsymbol{\varkappa}$ is the momentum of the absorbed photon. We have, with obvious notations :

$$(1s\,|\,Op\,|\exp\{i\mathbf{k}\,.\,\mathbf{r}\})=k_e(1s,\exp\{i(\mathbf{k}-\boldsymbol{\varkappa})\mathbf{r}\}).$$

With the use of this relation, one obtains easily the K-absorption coefficient near the edge (cf. Heitler 1936 b) :

$$K=4\sqrt{(2)}N\phi_0(137)^3\frac{Z_e^{5}E^{3/2}}{(\frac{1}{2}Z_e^{2}+E)^4h\nu}.$$

The definitions are the same as in (1), and momentum $\boldsymbol{\varkappa}$ has been neglected in comparison with $\mathbf{k}$.*

To take account of the fact that the bottom of the conduction band has a *negative* potential energy $-E_0$, we replace in K the energy change $(\frac{1}{2}Z_e^{2}+E)$ of the expelled electron by the more exact formula :

$$h\nu'=\tfrac{1}{2}Z_x^{2}+E,$$

where

$$Z_x^{2}=Z_e^{2}-2E_0$$

and thus write :

$$K=4\sqrt{(2)}N\phi_0(137)^3\frac{Z_x^{5}E^{3/2}}{(h\nu')^4h\nu}. \qquad (3)$$

A similar computation gives for the L and M absorptions :

$$\left.\begin{aligned}
L_I&=\frac{N\phi_0}{\sqrt{2}}(137)^3\frac{Z_x^{5}E^{3/2}(\frac{1}{8}Z_x^{2}-E)^2}{(h\nu')^6h\nu},\\
L_{II,III}&=\frac{3N\phi_0}{2\sqrt{2}}(137)^3\frac{Z_x^{7}E^{5/2}}{(h\nu')^6h\nu},\\
M_I&=\frac{N\phi_0}{486\sqrt{2}}(137)^3\frac{Z_x^{5}E^{3/2}[3(\frac{1}{3}Z_x)^4-20(\frac{1}{3}Z_x)^2E+12E^2]^2}{(h\nu')^8h\nu},\\
M_{II,III}&=\frac{12\sqrt{(2)}N\phi_0}{81}(137)^3\frac{Z_x^{7}E^{5/2}(\frac{5}{54}Z_x^{2}-E)^2}{(h\nu')^8h\nu},\\
M_{IV,V}&=\frac{N\phi_0}{405\times27\sqrt{2}}(137)^3\frac{Z_x^{7}E^{5/2}[3(\frac{1}{3}Z_x)^4-50(\frac{1}{3}Z_x)^2E+16E^2]^2}{(h\nu')^{10}h\nu},
\end{aligned}\right\} \qquad (4)$$

with $$h\nu'=\tfrac{1}{8}Z_x^{2}+E=\tfrac{1}{8}Z_e^{2}+E-E_0$$

for L absorption, and

$$h\nu'=\tfrac{1}{18}Z_x^{2}+E=\tfrac{1}{18}Z_e^{2}+E-E_0$$

for M absorption.

* Heitler's final formula is an approximation for large E ; this exact form is, however, easily deduced from his computations.

The absorption jumps computed from (3) and (4) are compared in table 1 with experimental values (Skinner and Johnston, *loc. cit.*; Johnston 1939, Meier 1910). We use Sommerfeld's value for the energy E at the top of the conduction band, and the 'effective charges' Z_e according to Slater's empirical rule (Slater 1930): Z_e is $Z-0{\cdot}3$ for K, $Z-4{\cdot}15$ for L and $Z-11{\cdot}25$ for M. Finally, E_0 is deduced from the cohesive energy E_s, the first ionization potential E_I and the number n of valency electrons by the formula (Mott and Jones 1936):

$$E_0 \simeq n^{-1}E_s + E_\mathrm{I} + 0{\cdot}6E.$$

Table 1. The Discontinuities τ of the Absorption Coefficient at the K, L, M edges

	$h\nu$ ev	E ev	E_0 ev	$\tau\times10^4$ calc.	$\tau\times10^4$ exp.
LiK	54·5	$4{\cdot}7_5$	10·0	1·35	3·0
MgL_I	63·0	7·3	12·7	2·5	0·07 or 1·8 ?
$MgL_{\mathrm{II,\,III}}$	49·2			1·4	1·4
$Al\,L_\mathrm{I}$	87·3	12·0	14·3	3·6	0·2 or 2 ?
$Al\,L_{\mathrm{II,\,III}}$	72·6			2·7	3·1
CuM_I	110·7	7·1	15·1	$1{\cdot}3_5$	$\simeq$0·5
$CuM_{\mathrm{II,\,III}}$	74·4			1·7	1·0
$CuM_{\mathrm{IV,\,V}}$	2·2			11·5	16·8
$ZnM_{\mathrm{II,\,III}}$	86·9	6·2	13·8	0·7	1·4

A correct *order of magnitude* is obtained down to the optical region ($Cu\,M_{\mathrm{IV,\,V}}$), except perhaps for L_I and M_I absorption. The interpretation of the experimental results for those edges is however not certain (cf. Johnston, *loc cit.*).

We show in Appendix (1) how in a better approximation the perturbing potential acting on the expelled electron could be taken into account. The correction introduced is small. But to interpret the fine structure of the absorption edge it is necessary to take into account the changes in the wave functions of the remaining electrons, both in the excited atom and in the conduction band. This is a more complex problem, which we shall now study in §§ 3 to 6. Only the simplest case, that of lithium, is analysed.

§3. Transitions in Many Electron-Systems

The final state of a metal after the absorption of a quantum of radiation is by hypothesis a stationary state, and the wave functions of all the electrons are of course slightly different from what they were in the initial state. We study in this section a simple formula which takes this change into account. It is based on determinantal wave functions, an approximation valid in the case of light elements where a definite

spin may be attributed to every electron. We shall apply this formula to Fermi electrons in §4, before using it in §§5 and 6 to study the fine structure.

The initial and final states of the system are thus represented by two orthonormal wave functions of the type :

$$\Psi_i=(N!)^{-1/2}\sum_{P_i}(-1)^{p_i}P_i\psi_{i1}(1)\psi_{i2}(2)\ .\ .\ .\ .\ \psi_{iN}(N),$$

$$\Psi_f=(N!)^{-1/2}\sum_{P_f}(-1)^{p_f}P_f\psi_{f1}(1)\psi_{f2}(2)\ .\ .\ .\ .\ \psi_{fN}(N),$$

where $p_j=1$ or 2 according to the parity of permutation P_j. The individual initial wave functions ψ_{ij} are assumed to be orthogonal to each other ; and similarly for the ψ_{fk}. But the ψ_{fk} are *not* assumed to be orthogonal to the ψ_{ij}.

The matrix element defined in §2 for the transition may be given a simple form (Pauling and Wilson 1935 a). We have

$$\begin{aligned}J&=\int\Psi_f^*Op\Psi_i\,d\tau\\&=(N!)^{-1}\sum_{P_f,P_i}(-1)^{p_f+p_i}P_fP_i\sum_{j=1}^{N}\int\psi_{f1}^*(1)\psi_{i1}(1)\,d\tau_1\ .\ .\ .\ .\\&\qquad\int\psi_{fj}^*(j)Op\psi_{ij}(j)\,d\tau_j\ .\ .\ .\ .\int\psi_{fN}^*(N)\psi_{iN}(N)\,d\tau_N.\end{aligned}$$

We call now P the permutation of the f's with respect to the i's :

$$P_f=PP_i.$$

J may be written :

$$\begin{aligned}J&=(N!)^{-1}\sum_{P}(-1)^{p}P\sum_{P_i}P_i\sum_{j=1}^{N}\int\psi_{f1}^*(1)\psi_{i1}(1)\,d\tau_1\ .\ .\ .\ .\\&\qquad\int\psi_{fj}^*(j)Op\psi_{ij}(j)\,d\tau_j\ .\ .\ .\ .\int\psi_{fN}^*(N)\psi_{iN}(N)\,d\tau_N.\end{aligned}$$

P_i only permutes the terms in the last summation over j without altering its value. As there are $N!$ such permutations, we have simply :

$$J=\sum_{P}(-1)^{p}P\sum_{j=1}^{N}\int\psi_{f1}^*\psi_{i1}\,d\tau\ .\ .\ .\ .\int\psi_{fj}^*Op\psi_{ij}\,d\tau\ .\ .\ .\ .\int\psi_{fN}^*\psi_{iN}\,d\tau.\quad .\quad (5)$$

To obtain J, we form therefore the determinant $|\int\psi_{fk}^*\psi_{ij}\,d\tau|$ of the projections between all the initial and final individual wave functions ; we then introduce operator Op in the terms of the jth row, and add the determinants obtained for $j=1$ to N.

Formula (5) may be interpreted as follows : the total transition probability is the sum of all the possible *transition* elements $\int\psi_{fk}^*Op\psi_{ij}\,d\tau$ between two individual wave functions, each one combined with all the possible *projection* elements $\int\psi_{fm}^*\psi_{in}\,d\tau$ of the remaining electrons. Each system of projection elements expresses a one to one correspondence between the initial and final wave functions, which may be pictured by arrows if the ψ_{fm} and ψ_{in} are placed on two parallel rows (cf. for instance, figs. 5, 6). The sign $(-1)^P$ of a projection is then given by the number of crossings of the arrows.

We stressed that the initial and final electronic wave functions ψ_{ij} and ψ_{fk} are not orthogonal. This is because in the Hartree–Fock approximation the central fields V and V' acting on them are different. We have for the two functions :

$$\nabla^2\psi_{ij}+(k^2-2V)\psi_{ij}=0,$$
$$\nabla^2\psi_{fk}+(k'^2-2V')\psi_{fk}=0.$$

The use of Green's theorem gives readily :

$$u_{jk}=\int\psi_{fk}^*\psi_{ij}\,d\tau=2(k'^2-k^2)^{-1}\int\psi_{fk}^*(V'-V)\psi_{ij}\,d\tau.$$

The projections may thus be considered as secondary transitions induced by the change of potential $\delta V=V'-V$ due to the main transition.

Hylleraas (1937) has treated the simplest application of formula (5), K absorption in helium : the remaining K electron must then be projected from a state of the full K shell to a hydrogenic wave function round a bare nucleus of charge 2. In metallic lithium, however, we must take into account the complex projections occurring in the conduction band. We devote the next section to this problem.

§4. Projections in a Fermi Gas

In this section, the initial and final states Ψ_i and Ψ_f are relative to an x-ray absorption process in a metal. Our purpose is to calculate some of the elements $\int\Psi_f^*\Psi_i\,d\tau$ which we call projection elements, when Ψ_i and Ψ_f are both entirely made up of conduction electrons and do not contain any bound electron. We obtain first an expression for the individual projection elements u_{mn} for one electron (§ 4.1, formula (9)). We then study in some detail the projection of a gas of many electrons (§§ 4.2 and 4.3). These elements will be used in §§ 5 and 6 to compute the total transition elements.

4.1. *Simple Form for the Individual Projections* u_{mn}

We give here an approximate form for the projections $u_{mn}=\int\psi_{fn}^*\psi_{im}\,d\tau$, where ψ_{im} and ψ_{fn} are one electron wave functions.

As this is an absorption process, the initial potential V is that of the lattice. We take it as a constant. The final potential $V'=V+\delta V$ differs from V by a certain amount δV, the difference between the potential of a normal ion and that of an ion excited by the absorption. In Li K absorption, for instance, δV is roughly the difference between the potentials of a Li^+ ion in the states $1s^2$ and 1s2s (cf. § 5).

We assume δV to be spherically symmetrical and take the metal to extend over a sphere of large radius R. The initial and final wave functions are thus of the type :

$$\psi_{im}=\frac{1}{2}\left(\frac{2l+1}{\pi R}\right)^{1/2}\frac{\phi_m(k,r)}{r}P_l(\cos\theta_m)\chi,$$
$$\psi_{fn}=\frac{A}{2}\left(\frac{2l+1}{\pi R}\right)^{1/2}\frac{\phi_n(k',r)}{r}P_l(\cos\theta_n)\chi,$$

where the spin wave functions χ and quantum numbers l must evidently be the same in the two functions. P_l is the Legendre coefficient; θ_m the angle of $\mathbf{r}$ with momentum $\mathbf{k}$. The radial parts satisfy:

$$\frac{d^2\phi_m}{dr^2}+\left(k^2-\frac{l(l+1)}{r^2}\right)\phi_m=0,$$

$$\frac{d^2\phi_n}{dr^2}+\left(k'^2-\frac{l(l+1)}{r^2}-2\delta V\right)\phi_n=0.$$

They are normalized so as to give for large r: $\phi_m \sim \sin(kr-\frac{1}{2}l\pi)$ and $\phi_n \sim \sin\{k'r+\eta_l(k')-\frac{1}{2}l\pi\}$, where η is the phase shift produced by δV.

A is a normalizing factor which takes into account the displacement of charge in the state of momentum k' due to the perturbing potential δV. If δV attracts locally a charge q, this reduces the density of charge outside the perturbed region by a factor $1-q$. One shows easily that $q=(1/R)\,[d\eta_l(k')/dk']$ (Friedel 1952 a). Thus:

$$A^2=1-\frac{1}{R}\frac{d\eta_l(k')}{dk'}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

There are $2l+1$ independent functions ψ_{im} of given l, and similarly for ψ_{fn}. Only the functions with parallel momenta $\mathbf{k}$ and $\mathbf{k}'$ have non-vanishing projections, so u_{mn} reduces to:

$$u_{mn}=4AR^{-1}(k'^2-k^2)^{-1}\int_0^R \phi_n(k'r)\delta V\phi_m(kr)\,dr. \quad . \quad . \quad . \quad (7)$$

Formula (7) is exact. We replace now the upper limit R of the integral by infinity. This adds a term, the magnitude of which oscillates as R increases, with an average value zero. This term may be neglected. Then developing $\phi_m(kr)$ and $\phi_n(k'r)$ successively in powers of $(k'-k)$, and using the exact relation (Mott and Massey 1949 a):

$$\int_0^\infty \phi_n(kr)\delta V\phi_m(kr)\,dr=-\tfrac{1}{2}k\sin\{\eta_l(k)\},$$

we have:

$$2\int_0^\infty \phi_n(k'r)\delta V\phi_m(kr)\,dr=-\tfrac{1}{2}[k'\sin\{\eta_l(k')\}+k\sin\{\eta_l(k)\}]+O(k'-k). \quad (8)$$

A more detailed discussion (cf. Appendix (2)) shows that this is a good approximation if k' is not too different from k and δV not too large.

The boundary condition $\psi_{fn}(k'R)=\psi_{im}(kR)=0$ gives:

$$kR\simeq\tfrac{1}{2}l\pi+m\pi \quad \text{and} \quad k'R\simeq\tfrac{1}{2}l\pi+n\pi-\eta_l(k'),$$

where m and n are integers which can be used to define the initial and final momenta. Thus:

$$u_{mn}\simeq\frac{A[k\sin\eta_l(k)+k'\sin\eta_l(k')]}{(k'+k)[\eta_l(k')+(m-n)\pi]}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where A is given by eqn. (6).

Formula (9) shows some general properties of the projection elements u_{mn}. First, if δV vanishes, u_{mn} becomes a δ function δ_{mn}, for free electrons have orthogonal wave functions. If now $\delta V\neq 0$, u_{mn} has,

for every m and n, a finite value, which is function only of the phase shifts η_l and does not depend otherwise on δV—this at least if $|\delta V|$ and $|k'-k|$ are not too large. The function $|u_{mn}|$ has a single maximum, for that value of $n-m$ which is nearest to η_l/π ; it tends to zero as $m-n$ tends to infinity.

When the metal becomes infinite (R infinite), u_{mn} will be negligibly small except for infinitely small values of $k'-k$. For such values :

$$u_{mn} \simeq \frac{\sin \eta_l(k)}{\eta_l(k)+(m-n)\pi}.$$

$|u_{mn}|$ is then a δ function if η_l is a multiple of π. But it has its broadest maximum when $d|u_{mn}|/dk=0$. This occurs when either $d\eta_l/dk=0$ or $\eta_l \simeq (\lambda+\frac{1}{2})\pi$ (λ integral).*

Finally, u_{mn} changes its sign, as $m-n$ varies, for $\eta_l=\lambda\pi$ and immediately after the maxima of $|u_{mn}|$.

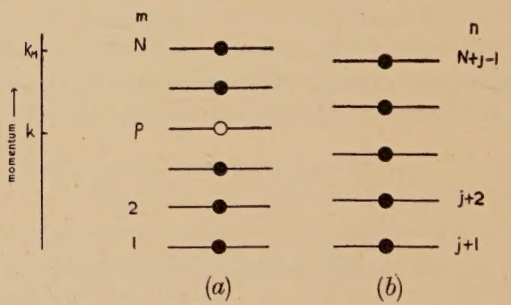
These remarks will be used in the next section. We study now the projection elements for a system of N Fermi electrons. We start with a case where the final state is the most stable possible.

4.2. *Fundamental Projections $I(p/N)$*

4.2.1. *General Description*

We restrict ourselves to a definite spherical component l and a definite spin, and study projections of a simple kind pictured in fig. 1. In the

Fig. 1



Scheme for projection $I(p/N)$. (a) Initial, and (b) final state.

initial state, $N-1$ electrons and a hole occupy the N states of lowest energy up to momentum k_M. In the final state, the $N-1$ electrons occupy the states of lowest energy *in the conduction band*. In fig. 1 (b) δV was assumed to subtract j bound states from the bottom of the band, so that the first state in the band is $n=j+1$. The total projection is denoted by $I(p/N)$ when the hole is in the pth state. Projections of this type will be used in § 5, the hole being replaced by an electron which, in the final state, drops into a bound level.

The quantity $(-1)^p I(p/N)$ is deduced from the matrix (u_{mn}) ($m=1$ to N, $n=j+1$ to $N+j-1$) by suppressing the $m=p$th row (§ 3).

* More exactly $tg[\eta_l+(m-n)\pi]=[\eta_l+(m-n)\pi]^2$.

Its large terms are near to the main diagonal, and its absolute magnitude is evidently smaller than one. Some general characteristics are easily deduced from the behaviour of u_{mn}.

Let $k=k_M(p+\frac{1}{2}l)/(N+\frac{1}{2}l)$ be the momentum of the hole p. If first $\eta_l(k)=\lambda\pi$ (λ integer), the projection $I_l(p/N)$ tends to zero as R tends to infinity. For, when R becomes very large, the projections of the $n=(p+\lambda)$th column tend towards a δ function (cf. §4.1):

$$u_{m,p+\lambda}\simeq\delta_{m,p}.$$

Therefore in the determinant $I_l(p/N)$ where $m\neq p$, all the terms of the $(p+\lambda)$th column tend to zero, and hence $I_l(p/N)\simeq 0$. As an application, $I_0(0)\simeq 0$, for at the bottom of the conduction band $\eta_0(0)=j\pi$ if there are j bound states (Mott and Massey 1949 b).

On the other hand, $dI_l/dk=0$ if $d|u_{mn}|/dk=0$ near the main diagonal. Hence (§4.1) either $d\eta_l(k)/dk=0$ or $\eta_l(k)\simeq(\lambda+\frac{1}{2})\pi$ (λ integral).

In conclusion, the projection is maximum if the momentum k of the hole is in a zone of maximum ' misfit ' with the quantized momenta k'. This happens if the momentum k falls just in the middle between two values of k' ($\eta_l=(\lambda+\frac{1}{2})\pi$) or if the misfit, hence the phase shift η_l, are stationary ($d\eta_l(k)/dk=0$). The projection on the contrary vanishes if the momentum k of the hole is just equal to one of the momenta k' ($\eta_l(k)=\lambda\pi$).

4.2.2. *Simple Example*

Keeping the band width k_M constant and allowing R to increase, each term in the development of $I(p/N)$ decreases, and tends to zero for an infinite volume of metal. But we believe that the number of terms in $I(p/N)$ must increase in such a way that the total tends to a finite value. This point is illustrated, if not proved, by the study of a simple example.

When δV has the form of a spherical well of radius a and constant depth $-D$, an exact and straightforward computation gives:

$$u_{mn}=\frac{2D}{R}\frac{X}{k^2-k'^2}\left[\frac{\sin(K'-k)a}{K'-k}-\frac{\sin(K'+k)a}{K'+k}\right]$$

with

$$X^{-2}=R^{-1}[a-(2k')^{-1}\sin 2K'a]+\sin^2 K'a\sin^{-2}(k'a+\eta_0)$$
$$[1-R^{-1}\{a-(2k')^{-1}\sin 2(k'a+\eta_0)\}],$$

$$k'\operatorname{cotg}(\eta_0+k'a)=K'\operatorname{cotg}K'a$$

and

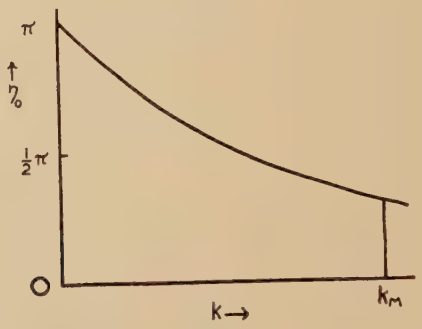
$$K'^2=k'^2+2D.$$

We took as an example (in atomic units): $a=3$ and $D=2/9$. There is then one bound s-state and $\eta_0(k)$ varies as pictured in fig. 2.

$I(p/N)$ was computed for $k_M=0{\cdot}63$ and $R=5$, 10, 15, 20 and 25, hence $N=1$, 2, 3, 4, and 5 (fig. 3).

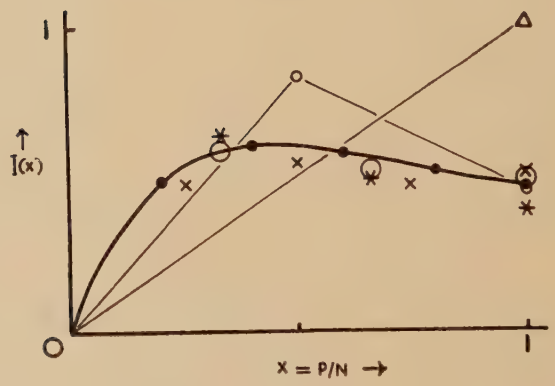
For $N \geqslant 3$, the curves seem to converge rapidly towards a *finite* limit, represented by the full curve. This convergence has been checked on other examples and seems reasonable, for there are already about 100 Li atoms in the sphere of radius $R=15$, and more than 450 for $N=5$. The small scatter of the points may be due to slight errors in the computation. As pointed out in §4.2.1, $I(0)=0$ and $dI/dx=0$ for $\eta_0=\pi/2$.

Fig. 2



$\eta_0(k)$ in a simple example.

Fig. 3



$I(p/N)$ for various N's.

Points marked △ for $N=1$; ○ for $N=2$; ∗ for $N=3$; × for $N=4$; ● for $N=5$; ○ for $N=3$ by approximation (9).

Failing a rigorous proof, we feel therefore justified in evaluating the projections $I(p/N)$ for small values of R($N=3$ or 4). The values obtained using approximation (9) are in fair agreement with the exact ones (fig. 3 for $N=3$). We shall therefore use this approximation in all further computations.

4.3. *Projections over Excited States*

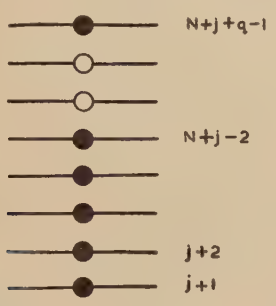
Projections of the initial state fig. 1 (a) over excited states will be used in §§5 and 6.

In the final state pictured in fig. 4, the electron with highest energy is separated from the rest by q holes. We denote by $I^q(p/N)$ such a projection. $I^q(1)$ is deduced from $I(1)$ by replacing the terms $u_{m,N+j-1}$ by $u_{m,N+j+q-1}$ which decrease rapidly as q increases. So only projections with small q are of importance. As for $I(1)$, one can show that, *for given* q, $I^q(1)$ tends rapidly towards a limit as the volume of the metal increases ; and it is sufficient to evaluate it for small R. $I^q(p/N)$ is then deduced by the approximate relation :

$$I^q(p/N) \simeq \{I^q(1)/I(1)\} I(p/N). \quad . \quad . \quad . \quad . \quad . \quad (10)$$

More complex excited states may be studied in the same way. They are of little importance in the transitions studied below.

Fig. 4



Final state for projection $I^q(p/N)$.

§5. Main Absorption Edge in Lithium

In a metal the bound states of the excited ion and the Fermi electrons are seriously affected by an x-ray absorption. They must all be taken into account in a detailed study of the transition ; and, according to formula (5), we must add the various matrix elements obtained when one electron makes a transition to any one of the new electronic states, all the other electrons being ' projected ' over the remaining states.

In lithium, three of these elements are preponderant : the transition may affect the inner bound electron ; this electron is then thrown into the continuum (fig. 5) or goes over to the new bound state which screens the hole created in the inner shell (fig. 16) ; the transition may also occur between a Fermi electron and the screening bound state (fig. 6). These transitions are appreciably large only when the Fermi electrons remain in their most stable state, thus when there is *no hole left in the conduction band*. The transition pictured fig. 16, where no electron is thrown into the continuum must correspond to an absorption line, while the two others give an absorption edge. The transition of fig. 16 is actually just an inner excitation of the monovalent ion, and only occurs when allowed by the usual selection rules.

5.1. *General Description*

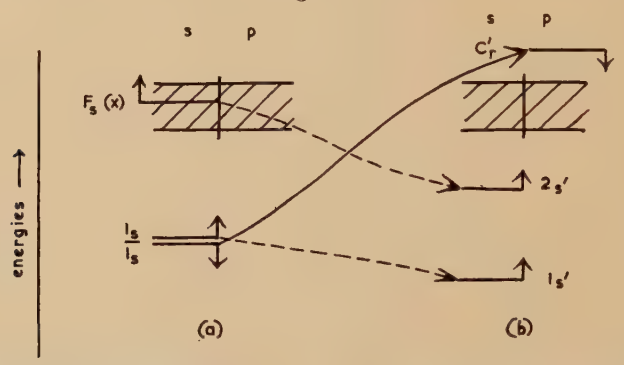
In lithium, the energetically most favourable screening of the hole in the K shell is by a 2s electron with spin parallel to that of the remaining K electron. For this configuration gives the most stable form of the lithium ion in the state 1s2s (Bacher and Goudsmit 1932).

We picture therefore the transition near the main absorption edge ($h\nu = 54{\cdot}5$ ev) in the following way. The initial state contains the K shell ($1s^2$) and n Fermi electrons (F_i). The final state contains a 1s and a 2s electrons with parallel spins (1s′, 2s′), one electron in the continuum (c') and $n-1$ Fermi electrons (F_i'). The dashes emphasize the fact that the potential provided by the excited ion is different from that of a normal ion ; hence the quantized states after absorption (1s′, F_i' . . .) are different from those before absorption (1s, F_i).

A possible scheme of transition is pictured in fig. 5. It may be written with an obvious notation :

$$T=(\overline{c_p}'|Op|\overline{1s})(1s', 1s)(2s', F_s(x))I_0(x)I_1{}^3(1)I_2{}^5(1)\ldots\overline{I_0}(1)\overline{I_1{}^3}(1)\overline{I_2{}^5}(1).\ldots$$

Fig. 5



First schema of absorption with 1s′2s′. (*a*) Initial and (*b*) final states : ——→ transition ; – – → projection.

The terms barred are those with spin antiparallel to that of 2s′ and $F_s(x)$ denotes a Fermi state of s symmetry with momentum $k=xk_M$. The factor in I_l is obtained assuming that at the end of the transition the Fermi electrons are in their most stable state (cf. §4.2), i.e. there is *no hole left in the Fermi gas at the end of the absorption.*

Such transitions, whatever the value of x, are between states of the same difference of energy. The total transition probability is obtained by integrating T^*T over x after multiplication by the density of states ρ_x of the Fermi electrons $F_s(x)$: the Fermi electron which happens to be nearest to the atom during the absorption drops into the 2s′ state ; its energy may have any value in the range of the conduction band.*

* As $I_0(0)=0$, the electron cannot come from the bottom of the conduction band. We shall see that the main contribution comes usually from the top of the band (cf. fig. 10).

The total transition probability is therefore finite, although each individual T tends to zero if the volume of the metal becomes infinite. Transitions of the type pictured in fig. 5 therefore correspond to a finite absorption edge at the critical frequency.

There are many possible schemes of absorption other than T. We shall however neglect those where the electronic transitions are between Fermi electrons, which are certainly unimportant. Conservation of spin and usual selection rules for electronic transitions then restrict the possibilities. The transitions leading to the most stable states, i.e. those which leave no hole in the Fermi gas, add up to two terms corresponding to two different configurations :

$$\begin{aligned} A(x)=&(\overline{c_p'}|Op|\overline{1s})[(1s', 1s)(2s', F_s(x)) \\ &-(1s', F_s(x))(2s', 1s)]I_0(x)I_1{}^3(1)\ldots\overline{I_0}(1)\ldots \\ B(x\,;y)=&\{(2s'|Op|F_p(y))[(1s', 1s)(\overline{c_s'}, \overline{1s})I_0(1)-(\overline{c_s'}, \overline{1s})(1s', F_s(x))I_0(x)] \\ &-(1s'|Op|F_p(y))(\overline{c_s'}, \overline{1s})[(2s', 1s)I_0(1) \\ &-(2s', F_s(x))(F_s'(1), 1s)I_0(x)]\}I_1(y)I_1{}^2(1)I_2{}^5(1)\ldots\overline{I_0}(1).\ldots \end{aligned}$$

The various signs are easily obtained by counting the number of crossings of the transition arrows (§3). The total transition probability is obtained by integration over x and y :

$$\int_0^1 A^*A\rho_x\,dx+\int_0^1\int_0^1 B^*B\rho_x\rho_y\,dx\,dy. \quad . \quad . \quad . \quad . \quad (11)$$

Transitions leading to holes in the Fermi gas will be considered in §5.4. But we shall first study the main term (11).

5.2. *Reduction of Formula* (11)

We use spherical harmonics for the states in the conduction band and the continuum, and hydrogenic wave functions for the bound states (cf. (2) for 1s).

Thus :

$$\left.\begin{aligned} F_s(x)&=V^{-1/2}k^{-1}r^{-1}\sin kr, \\ c_p'&=V^{-1/2}3ik'^{-2}r^{-2}\,(\sin k'r-k'r\cos k'r)P_1(\cos\theta), \\ 2s'&=(4\sqrt{(2\pi)})^{-1}Z'_e{}^{3/2}(2-Z_e'r)\exp(-\tfrac{1}{2}Z'_er). \end{aligned}\right\} \quad (12)$$

We have then :

$$\left.\begin{aligned} (1s', 1s)&=8(Z_e+Z)^{-3}(Z_eZ)^{3/2}, \\ (2s', 1s)&=2\sqrt{(2)}(Z_e+\tfrac{1}{2}Z'_e)^{-4}(Z_eZ'_e)^{3/2}(Z_e-Z'_e), \\ [2s', F_s(x)]&=-V^{-1/2}2\sqrt{(2\pi)}(\tfrac{1}{4}Z'_e{}^2+k^2)^{-3}Z'_e{}^{5/2}(\tfrac{1}{4}Z'_e{}^2-k^2). \end{aligned}\right\} \quad (13)$$

As in §2, we take into account the negative energy $-E_0$ of the bottom of the conduction band by replacing in this last projection Z'_e by Z'_x chosen so that $\frac{1}{8}Z'_x{}^2=\frac{1}{8}Z'_e{}^2-E_0$.

We may attribute to the screening electron 2s′ the difference of energy between the absorption edge $h\nu_0$ and the second ionization potential of lithium Li II (cf. Friedel 1952 a) :

$$E_{2s'} \simeq \text{Li II} - h\nu_0.$$

We have $h\nu_0 = 54{\cdot}5$ ev (Skinner and Johnston *loc. cit.*) ; Li II $= 75{\cdot}26$ ev (Bacher and Goudsmit *loc. cit.*). Hence $E_{2s'} = \frac{1}{8}Z'_e{}^2 = 20{\cdot}8$ ev and $Z'_e \simeq 2{\cdot}4_8$.

(13) then gives $(1s', 1s) = 0{\cdot}997_5$, $(2s', 1s) = 0{\cdot}044_5$ and in (11) only the terms containing (1s′, 1s) are of importance. The absorption coefficient thus reduces to :

$$\tau_1 \simeq K I_1{}^4(1) I_2{}^{10}(1) \ldots \overline{I_0}{}^2(1) \ldots (1s', 1s)^2 \int_0^1 (2s', F_s(x))^2$$
$$[I_0{}^2(x) I_1{}^2(1) + I_0{}^2(1) I_1{}^2(x)] \rho_x \, dx \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

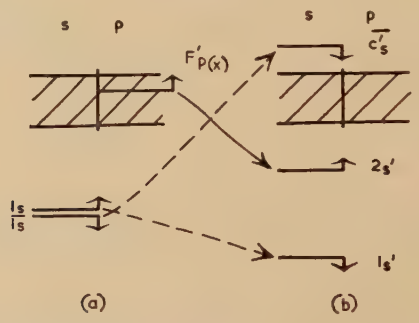
with

$$\rho_x \, dx = (2\pi^2)^{-1} V k^2 \, dk.$$

K is given by (3) and the various projections by (13).

The two terms in τ_1 are described in figs. 5 and 6.

Fig. 6



Second scheme of absorption with 1s′2s′ : (*a*) initial and (*b*) final states. ——→ transition ; – – → projections.

5.3. *Jump in Absorption at the Edge*

We shall now compute (14) when the electron thrown into the continuum is just at the top of the conduction band : $k' = k_M$.

We have to evaluate the projection factors I_l, hence the phase shifts η_l produced in the Fermi gas by the perturbing potential δV. The main part in δV is due to the replacement of a normal Li^+1s^2 ion by an excited 1s′2s′ one. We shall neglect the other terms, namely the contribution from the charge displaced in the conduction band and the corrective term in Wannier's theorem (cf. Friedel 1952 a). These should be considered to obtain a more correct self-consistent value of δV.

We divide δV into two parts, an electrostatic δV_c and one due to exchange δV_e :

$$\delta V = \delta V_c + \delta V_e.$$

With (2) and (12) we have :

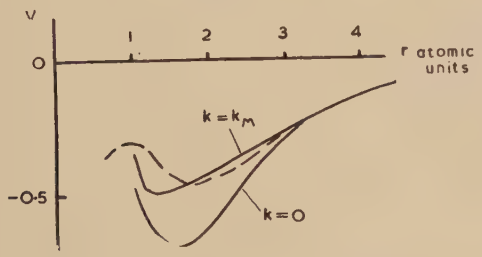
$$\delta V_c = r^{-1}[2(1+Z_e r)\exp(-2Z_e r) - (1+Zr)\exp(-2Zr) - (1+\tfrac{3}{4}Z_e' r + \tfrac{1}{4}Z'^2_e r^2 + \tfrac{1}{8}Z'^3_e r^3)\exp(-Z'_e r)].$$

For δV_e, we used Slater's average value (1951) :

$$\delta V_e \simeq \frac{1{\cdot}272}{r^2 \delta V_c}\left[r^2 \frac{d^2(r\delta V_c)}{dr^2}\right]^{1/3} \quad \ldots\ldots \quad (15)$$

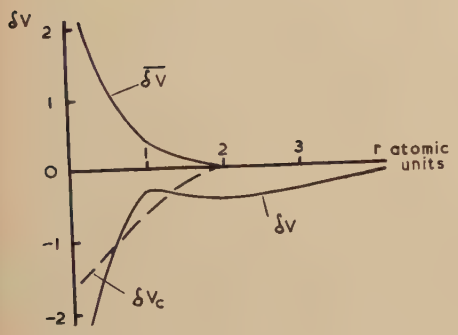
We checked this formula by a direct evaluation of the exchange term between the 2s′ electron and the Fermi electrons with s symmetry (cf. Fock and Petrashen 1935). Figure 7 shows a fair agreement in

Fig. 7



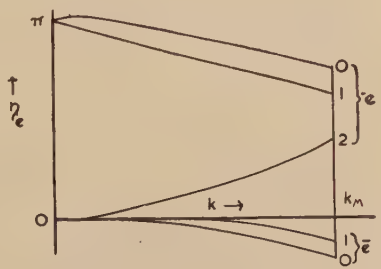
Term in δV_e due to electron 2s′. —— Fock ; – – – Slater.

Fig. 8



Perturbing potentials for 1s′ 2s′.

Fig. 9



Phase shifts for 1s′ 2s′.

the interesting range of radii ($r>2$ for $k \leqslant k_M = 0{\cdot}59$). Thus for Fermi electrons with spin parallel to that of 2s′ :

$$\delta V_e = 1{\cdot}272[2^{2/3}Z_e \exp(-\tfrac{2}{3}Z_e r) - \{4Z^3 \exp(-2Zr) + \tfrac{1}{2}Z'^2_e(\tfrac{1}{2} - Z'_e r + \tfrac{1}{4}Z'^2_e r^2)\exp(-Z'_e r)\}^{1/3}]$$

and, for Fermi electrons with opposite spin :

$$\overline{\delta V_e} = 1{\cdot}272 \times 2^{2/3} Z_e \exp(-\tfrac{2}{3}Z_e r).$$

δV and $\overline{\delta V}$ are given fig. 8. Owing to exchange with 2s′, the potential δV is attractive enough to subtract from the conduction band

one s and one p state with spins parallel to its own : $\eta_0(0)=\eta_1(0)=\pi$ (fig. 9). The $\overline{\eta_l}$ are small. As our δV and $\overline{\delta V}$ are not self-consistent, the η_l do not exactly satisfy the screening condition (Friedel 1952). We have actually :

$$\pi^{-1}\sum_l(2l+1)[\eta_l(k_M)+\overline{\eta_l}(k_M)]-4=-1{\cdot}1\neq 0.$$

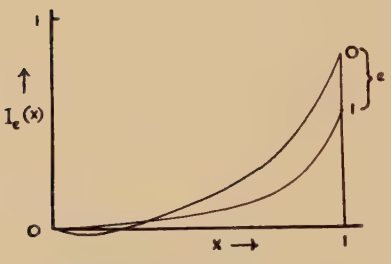
From η_l we deduce $I_l(x)$ as described in § 4.2. Figure 10 gives I_0 and I_1. The others are practically 1 for $x=1$ and 0 elsewhere.

With these values for the I_l, relation (14) gives :

$$\tau_1=(0{\cdot}04+0{\cdot}04_5)\times K=0{\cdot}11_5\times 10^4.$$

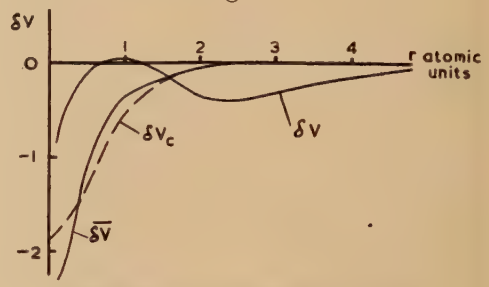
The two terms correspond to the schemes of figs. 5 and 6. They are of the same order of magnitude : it is here equally easy to subtract an electron from the s or p part of the conduction band, because the corresponding phase shifts η_0 and η_1, hence the projection factors I_0 and I_1, are similar. We see also that the major contribution of the Fermi electrons comes from the top of the conduction band.

Fig. 10



Projections $I_l(x)$ for 1s′ 2s′.

Fig. 11



Perturbing potentials for $\overline{1s'}$ 2s′.

5.4. *Variation of τ with the Energy of Absorption*

τ_1 represents the two major terms in (11). The others contribute to only a few per cent of τ_1, as is easily computed, and may be neglected.

But to every ' fundamental ' transition so far studied there corresponds a series of transitions of the type described in § 4.3 leading to an excited state of the Fermi gas. Here, the only ones of importance are the $I^q(x)$ with q small ($q<q_0=3$ or 4). So on the short wave lengths side of the edge, the absorption coefficient increases rapidly from τ_1 to a saturation value τ. The change occurs within a finite number of Fermi states, and hence in an infinitesimal range of frequencies ; and the coefficient experimentally measured is τ. Using relation (10), we must therefore replace every projection $I^2(x)$ by

$$I^2(x)+\sum_q[I^q(x)]^2\simeq I^2(x)\left(1+\sum_q\left[\frac{I^q(1)}{I(1)}\right]^2\right).$$

This is greater by 1·12 for I_0^2 and $1{\cdot}12_5$ for I_1^2 ; thus τ is increased by a factor 1·6 :

$$\tau\simeq 1{\cdot}6\tau_1=0{\cdot}2\times 10^4.$$

Finally, τ varies with the energy of absorption proportionally to K. Thus from (3)

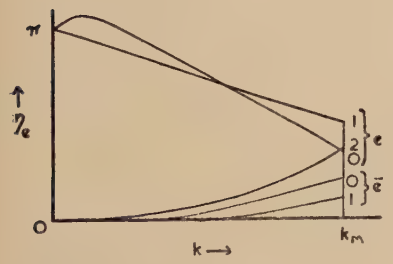
$$\frac{d\tau}{\tau d(h\nu)} = \frac{3}{2E} - \frac{4}{h\nu^1} - \frac{1}{h\nu} = 0{\cdot}0_4 \text{ ev}^{-1}.$$

τ should therefore increase slightly on the short wave lengths side of the edge, up to a maximum at a distance from the edge found to be about 25 ev. However, before this energy is reached, other features occur, probably due to screening by excited electrons.

§6. Screening by Excited Bound States in Lithium

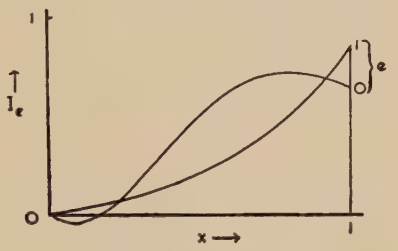
In a previous paper (Friedel 1952 a) we interpreted tentatively the fine structure on the short wave lengths side of the main absorption edge as due to screening by more excited electrons. We show in this section that one should expect an absorption *edge* for every configuration of the inner ion ($1s'\overline{2s}'$, $1s'2s'$, $1s'\overline{2p}'$, $1s'2p'$ . . .) and an absorption *line* when a direct excitation of the ion is possible ($\overline{1s}1s \rightarrow \overline{1s}'2p'$ . . . .).

Fig. 12



Phase shifts for $\overline{1s}'$ 2s'.

Fig. 13



Projections for $\overline{1s}'$ 2s'.

6.1. *Secondary Absorption Edges*

The same analysis as for the main edge gives the following results.

6.1.1. $\overline{1s}'2s'$

The energy difference between the $\overline{1s}'2s'$ and $1s'2s'$ states of Li^+ ion is about 1·5 ev (from optical data, cf. Bacher and Goudsmit *loc. cit.*). Attributing to electron 2s' this difference, we have $\frac{1}{8}Z'_e{}^2 = E_{2s'} \simeq 20{\cdot}8 - 1{\cdot}5 = 19{\cdot}3$ ev and $Z'_e = 2{\cdot}38$. The coulomb potential δV_c is therefore slightly more attractive. The exchange term due to 1s' is now added to $\overline{\delta V}$ and subtracted from δV (fig. 11). So the $\overline{\eta_l}$ are positive and η_0 smaller for large k (fig. 12), and I_0 has a maximum for $\eta_0 = \frac{1}{2}\pi$ (cf. §4.2). I_0 and I_1 are given in fig. 13. $I_2(1) \simeq 0{\cdot}98$ and the other $I(1)$ may be taken as 1.

The possible transitions are, with a change of spins, those pictured in figs. 5 and 6. And (14) gives :

$$\tau_1 \simeq (0{\cdot}11 + 0{\cdot}04) \times 10^4 = 0{\cdot}15 \times 10^4.$$

Projection over excited states (§ 5.4) multiply τ_1 by a factor 1·7. Thus :

$$\tau \simeq 1{\cdot}7\tau_1 = 0{\cdot}25_5 \times 10^4.$$

6.1.2. $\overline{1s}'2p'$

We take for 2p′ a hydrogenic wave function :

$$2p' = [4\sqrt{(2\pi)}]^{-1} Z''_e{}^{5/2} r \exp(-\tfrac{1}{2} Z''_e r) \cos\theta \quad . \quad . \quad . \quad (16)$$

where the value of Z''_e is deduced from the position of the absorption line L_1 (fig. 17) which we interpret as corresponding to the same screening $\overline{1s}'2p'$. We have $h\nu = 58{\cdot}5$ ev (Skinner and Johnston *loc. cit.*), hence $Z''_e \simeq 2{\cdot}22$. δV, η_l, I_0 and I_1 are very similar to those for 1s′2s′, and $I_2(1) = 0{\cdot}9$.

Thus practically only Fermi electrons with s or p symmetry can drop into the bound state. And the absorption is the sum of two transitions similar to those pictured in figs. 5 and 6, with a permutation of s and p components in the Fermi gas :

$$\tau_1 = 3K I_1{}^4(1) I_2{}^{10}(1) \ldots \overline{I_0{}^2}(1) \ldots (1s'1s)^2$$
$$\int_0^1 |\, 2p', F_p(x)\,|^2 (I_0{}^2(x) I_1{}^2(1) + I_0{}^2(1) I_1{}^2(x)) \rho_x \, dx \quad . \quad . \quad (17)$$

where factor 3 is due to degenerescence of 2p′ states. With (12) and (16) one finds :

$$(2p', F_p(x))^2 = 8\pi (3V)^{-1} Z''_x{}^2 k^2 (\tfrac{1}{4} Z''_x{}^2 + k^2)^{-6}$$

where Z''_e has been replaced by Z''_x such that $\frac{1}{8} Z''_x{}^2 = \frac{1}{8} Z''_e{}^2 - E_0$ (cf. § 2).

Finally :

$$\tau_1 \simeq (0{\cdot}03 + 0{\cdot}05) 10^4 = 0{\cdot}08 \times 10^4.$$

Projections over excited states give a factor 1·5. Thus

$$\tau \simeq 1{\cdot}5\tau_1 = 0{\cdot}12 \times 10^4.$$

6.1.3. 1s′2p′

We neglect the difference of energy between this case and the preceding one (less than 1 ev from optical data), and thus use the same value for Z''_e. δV is then similar to that for 1s′2s′ but stronger. Hence $\eta_0(0) = 2\pi$ (fig. 14) ; I_0 presents a maximum for small k and is nearly zero at the top of the conduction band (fig. 15).

The same formula (17) as before is valid for τ_1 and gives :

$$\tau_1 \simeq (0{\cdot}000_5 + 0{\cdot}01_2) 10^4 = 0{\cdot}01_3 \times 10^4.$$

With a factor 1·5 for excited states, the total absorption coefficient τ is small :

$$\tau \simeq 1{\cdot}5\tau_1 = 0{\cdot}02 \times 10^4.$$

6.2. 1s′2p′ *Absorption Line*

With 1s′2p′ a direct transition of the inner Li^+ ion is possible (fig. 16) :

$$(\Psi_f{}^* |\, Op \,| \Psi_i) = (2p' |\, Op \,| 1s)(\overline{1s'}, \overline{1s}) I_0(1) I_1{}^3(1) \ldots . \overline{I_0}(1) \ldots .$$

There is no electron c' thrown into the continuum ; so this gives an absorption *line*.

The maximum strength of the line is given by (Heitler 1936 c) :

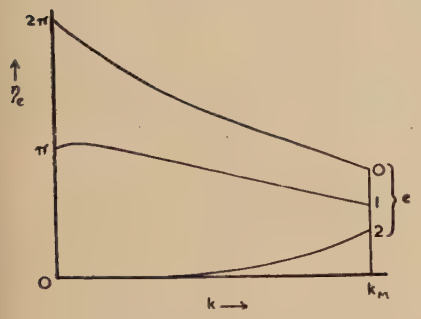
$$\tau_M = 2\pi\,(137)^2 \frac{N\hbar w_{ab} a_0^2}{(h\nu)^2 h\gamma}. \quad \ldots\ldots\ldots \quad (18)$$

N is the number of atoms per cm^3 ; $h\nu$ the resonance energy and $h\gamma$ the line width (both expressed in atomic units) ; and w_{ab} the transition probability, proportional to $h\nu |(\Psi_f^*| Op\, |\Psi_i)|^2$.

With (2) and (16), we have

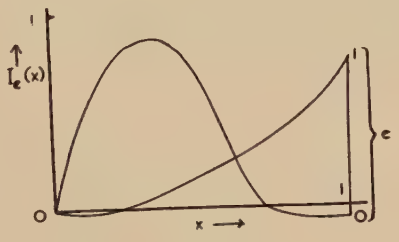
$$\hbar w_{ab} = 6\,\frac{8h\nu}{3\times(137)^2}\,\frac{(Z_e Z''_e)^5}{(Z_e+\frac{1}{2}Z''_e)^8}\,(1s', 1s)^2\, I_0^2(1)\, I_1^6(1) \ldots \overline{I_0^2}(1) \ldots$$

Fig. 14



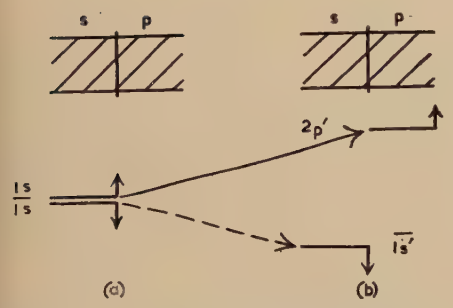
Phase shifts for 1s′2p′.

Fig. 15



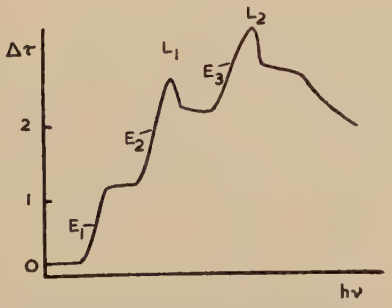
Projections I_l for 1s′2p′.

Fig. 16



Schema for absorption line with 1s′2p′. (*a*) Initial and (*b*) final state.

Fig. 17



Li K absorption, after Skinner and Johnston.

The factor 6 comes from the two K electrons and (2×3) 2p′ states. An experimental line width $\gamma \simeq 1{\cdot}5$ ev (Skinner and Johnston *loc. cit.*) and the values obtained in § 6.1.2 give

$$w_{ab} = 0{\cdot}45\times 10^{10}\ \text{CGS}.$$

Hence

$$\tau_{M_1} \simeq 1{\cdot}2\times 10^4.$$

Projections over more excited states in the Fermi gas (cf. § 4·3) introduce a factor 1·5. Thus finally

$$\tau_M \simeq 1{\cdot}8 \times 10^4.$$

The *large broadening* γ may be due in part to Auger effect, the 2p′ bound electron making a transition to the neighbouring 2s′ state when colliding with a conduction electron. The same broadening should affect the 1s′2p′ and 1s′2p′ edges. Some of γ may also be due to lack of resolving power in the instruments ; the form of the line is actually different in the two measurements by Skinner and Johnston.

§ 7. Comparison with Experiment

Skinner and Johnston's measurements (*loc. cit.*) give three successive absorption edges E_1, E_2 and E_3, E_2 and E_3 starting with somewhat broad and weak lines L_1 and L_2 (fig. 17).

E_1 probably corresponds to the formation of 1s′2s′ Li^+ ion, and the computed value of τ turns out somewhat too small (table 2).

Table 2. Comparison of the Fine Structure with Experiment

$\triangle\tau \times 10^{-4}$	E_1	E_2	L_1
exp.	1·0	1·2	0·5
comp.	0·2	0·4	1·8

The same happens with E_2, if we attribute it to the 1s′2s′ and broadened $\overline{1s}$′2p′ and 1s′2p′ edges. The computed value for L_1 on the other hand is too large. We believe this is because the 2p′ wave function is actually somewhat swollen and deformed by the surrounding lattice. For E_3 and L_2, the approximations concerning both bound and free electrons would certainly be unsatisfactory. Finally the absorption seems more or less constant after each edge, as predicted in § 5.4.

The computation gives therefore an acceptable order of magnitude, *but no very good agreement*, the calculated relative value of the L_1 line with respect to the edges being about ten times too strong. (cf. table 2).

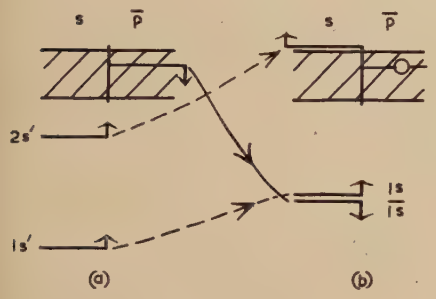
The fine structure E_2, L_1, E_3, L_2 . . . could on the other hand be a Kronig structure (Jones and Mott 1937). The first peak L_1 would then correspond to the peak in the first zone arising from reflection on the 110 planes. As pointed out by these authors, this peak would be observed in a K absorption only if the first zone is ' inverted ' so as to have s character at the bottom but p character at the top. In that case the K emission should have a sharp maximum at the edge. But actually no such maximum is observed (Skinner 1940). Also a fine structure similar to that of lithium, with same position of the secondary

maxima, occurs for LiOH which has of course another crystal structure (Skinner and Johnston, *loc. cit.*). These two facts seem to suggest that the structure observed is *not* a Kronig structure.

§8. Qualitative Description of Li K Emission

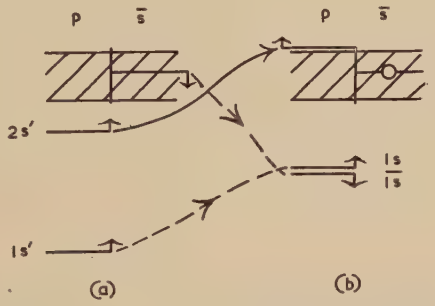
The transitions occurring in Li K emission are more or less the reverse of those for absorption. We point out some features of the process. The initial state in the emission contains mainly the most stable Li^+ ion, 1s′2s′, and the main transitions for the emission band correspond to those of figs. 5 and 6 for the absorption. They are pictured in figs. 18 and 19.

Fig. 18



First type of transition for the emission band. (*a*) Initial and (*b*) final state.

Fig. 19



Second type of transition for the emission band. (*a*) Initial and (*b*) final state.

An electron drops from the part of the conduction band with spin antiparallel to that of 2s′. This part is little perturbed ($\overline{\delta V}$ and $\overline{\eta_e}$ small in figs. 8 and 9) and the projections $I_l(x)$ are zero for $x \neq 1$ (§5.3). Conversely the perturbed Fermi gas with a $\overline{p}$ (or $\overline{s}$) hole of momentum $k < k_M$ must have a negligible projection over the stable unperturbed gas. A $\overline{p}$ (or $\overline{s}$) hole must therefore be created, with momentum practically equal to that k of the electron dropping into the 1s level. As for absorption, the process leaving a p hole (fig. 18) is probably the more important. s holes can nevertheless be created (fig. 19).

§9. Conclusion

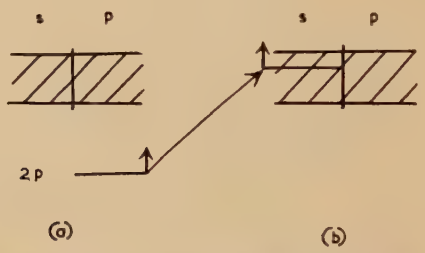
Possibility of Lines in the Fine Structure

In conclusion, experimental results seem compatible with a screening of the hole created in the K shell by absorption. In lithium, an absorption edge should correspond to every possible configuration of the screening and a line to every possible direct excitation of the inner ion Li^+. The computed strength of the absorption jump at the edges agrees in order of magnitude with experiment. It is smaller for more excited screenings, because the Fermi electrons are more perturbed.

Thus the edge computed for a 2p′ screening is smaller than for a 2s′ one ; for 1s′2p′, smaller than for $\overline{1}$s′2p′. In our interpretation of Johnston and Skinner's measurements, the 2p′ line is weaker than as computed, probably because the 2p′ state is somewhat swollen and deformed by the surrounding lattice.

Similar absorption lines should appear in various transitions of monovalent metals. At the edge of $L_{II, III}$ absorption, a weak line may actually be seen in Na (O'Bryan 1940), but not in Cu (Coster and Veldkamp 1932) nor in Au (Veldkamp 1935). We think that the line is not observed with heavy elements for the following reason : as the atomic number increases, the radius of the outer screening electron increases. The overlap on it by the inner L electron decreases therefore more quickly than the overlap on the states in the conduction band, which keep a nearly constant density near the origin. This decreases the importance of the lines (transitions of the type pictured in fig. 16) relative to the edges (cf. figs. 5 and 6). The rapidity of this decrease may be deduced from a remark by Bethe (1933 b) that, for a given atom

Fig. 20



$L_{II, III}$ line when screening is by Fermi electrons. (*a*) Initial and (*b*) final states.

or ion, the relative intensity of say transitions 2p—ns with various n are nearly the same as for a hydrogen atom. Thus the ratio of the oscillator strengths of transitions 2p—3s, 4s, 5s, 6s to that for transitions towards the continuum (a constant in Bethe's case) are proportional to 1, 0·2, 0·08 and 0·04.

As in lithium, the intensity of these absorption lines is also probably reduced by perturbation of the ns bound state by the surrounding lattice, especially in noble metals with their large d shells. Furthermore, the screening electron is probably replaced in noble metals by a heaping up of charge in the conduction band (Friedel 1952 a). In other words, the perturbing potential δV is perhaps not strong enough to subtract a bound state from the bottom of the conduction band, but increases only the density of Fermi states in k space, mainly in the s component. Transitions of the type pictured in fig. 16 are then impossible, and one may wonder whether any absorption line may be observed in those conditions. Actually a transition of the type pictured in fig. 20 may now take place ; owing to the crowding of the Fermi s states, the projection of the Fermi electrons pictured in fig. 20 may have a finite value. It will

correspond to an absorption *line*, for projections over more excited states of the Fermi gas will be vanishingly small. By integration over all possible final states, one obtains a transition probability with a finite value, which we believe to be of the same order of magnitude as if screening was by a bound state. If this is so, x-ray evidence cannot help in differentiating the two types of screening. Thus for instance we interpreted the secondary peak observed at about 2500 Å in optical absorption of copper as a direct 3d–4p excitation (Friedel 1952 b), but the 4p screening is certainly obtained by a heaping up of Fermi electrons and not by a bound state. We think therefore that it is only because of the weakness of the transition 2p–4s that no line is observed at the edge of Cu $L_{II, III}$.

Finally, in Na $L_{II, III}$ absorption, the line should correspond to the less stable singlet state L_{III} (configuration $^1P_1$ of Na^+), as is indeed observed. In Na $L_{II, III}$ emission, the main contribution to the band is thus the more stable triplet state L_{II} (configurations $^3P_2$, $^3P_1$, $^3P_0$ of Na^+), for which no direct transition of the Na^+ ion is possible. The contribution by L_{III} should contain a line at the emission edge. But this term is very weak in Mg (Skinner 1940) ; and actually no line has been observed either for Na $L_{II, III}$ (Skinner *loc. cit.*) or for the analogous K $M_{II, III}$ emissions (Kingston 1951).

The author wishes to thank Professor N. F. Mott for enlightening discussion and advice and the Direction des Mines for allowing him to stay in Bristol.

APPENDIX

(1) Approximate Correction for the Expelled Electron

In computing K (§ 2), we treated the ejected electron c' as a plane wave. We indicate here how one could take somewhat into account the perturbing potential δV acting on it.

We use a simple model where δV is a spherical well of potential of radius a and constant depth, at the centre of a large sphere of radius R representing the metal. For an s state with momentum k and if $ka > \pi$, the average radial density of charge may be considered as constant both inside and outside the well. When $\delta V = 0$, the charge of the state inside the well is thus

$$q \simeq a R^{-1}.$$

Let $q + \Delta q$ be the charge of the state inside the well when $\delta V \neq 0$. The radial density of charge of the state outside the well is now multiplied by $(1-q-\Delta q)(1-q)^{-1}$. The density of states per energy unit is multiplied by

$$\left(1 - \frac{1}{R}\frac{d\eta}{dk}\right)^{-1}$$

(Friedel 1952 a). And the product of the two, that is the total density of charge per energy unit outside the well, must remain constant (Laue 1914). Thus

$$1-q-\Delta q \simeq (1-q)(1-R^{-1}\, d\eta/dk)$$

or

$$(q+\Delta q)\, q^{-1} \simeq 1+(q^{-1}-1)\, R^{-1}\, d\eta/dk.$$

The total density of charge per energy unit inside the well is then multiplied by a factor

$$f \simeq [1+(a^{-1}-R^{-1})\, d\eta/dk][1-R^{-1}\, d\eta/dk]$$

or, for large R

$$f \simeq 1+\frac{1}{a}\frac{d\eta}{dk}.$$

If δV has a more general form, we take for a the radius giving, in the W K B approximation, the major contribution to the phase shift η. For the δV obtained in this paper for lithium, f is then always *slightly smaller than unity*, and therefore reduces τ somewhat.

(2) Form of the Projection Elements u_{mn} when $|k'-k|$ is large

The development (8), carried one step further, gives

$$O(k'-k)=\frac{k'-k}{k}\int_0^\infty r\delta V\left[\frac{\partial\phi_n(k,r)}{\partial r}\phi_m(k,r) - \phi_n(k,r)\frac{\partial\phi_m(k,r)}{\partial r}\right]dr+O(k'-k)^2.$$

A linear combination of the Schroedinger equations for ϕ_m and ϕ_n (cf. § 4.1) gives easily

$$O(k'-k)=2\frac{k'-k}{k}\int_0^\infty r\delta V(r)\, dr \int_0^r \delta V(r')\phi_m(k,r')\phi_n(k,r')\, dr'+O(k'-k)^2.$$

In the cases considered in this paper, δV is large only in a region near the origin where ϕ_m and ϕ_n have no zero. An *upper limit* for the absolute magnitude of $O(k'-k)$ is then obtained by replacing the limit of integration r by infinity. The second integral is then equal to $-\frac{1}{2}k \sin\{\eta(k)\}$ (cf. § 4.1), and

$$O(k'-k) \simeq -(k'-k) \sin\{\eta(k)\}\int_0^\infty r\delta V\, dr.$$

The maximum of $|k'-k|$, given by the maximum momentum k_M, is a fraction of atomic unit in monovalent metals. This development is therefore justified, and $O(k'-k)$ may be neglected, provided $\int_0^\infty r\delta V\, dr$ is not too large.

For very large values of $|k'-k|$ on the other hand, one of the functions ϕ_m or ϕ_n is rapidly oscillating and may be replaced by a sinus function. Application of a lemma by Riemann and Lebesgue (Whittaker and Watson 1950) shows that u_{mn} is of the order of

$$k^{-1}\int_0^\infty \delta V\, dr, \qquad k'^{-1}\int_0^\infty \delta V\, dr \qquad \text{or } (kk')^{-1}\int_0^\infty \delta V\, dr,$$

according respectively to whether k tends to infinity, k' remaining finite ; k' tends to infinity, k remaining finite or k, k' and $|k-k'|$ tend all three to infinity.

References

Bacher, R., and Goudsmit, S., 1932, *Atomic Energy States* (London).
Bethe, H., 1933 a, *Handbuch d. Phys. XXIV.* 1, 477 ; 1933 b, *Ibid.*, 469.
Coster, D., and Veldkamp, J., 1932, *Zeits. f. Phys.*, **74,** 191.
Fock, V., and Petrashen, M., 1935, *Phys. Zeits. Sowj.*, **8,** 547.
Friedel, J., 1952 a, *Phil. Mag.*, **43,** 153 ; 1952 b, *Proc. Phys. Soc.*, in press.
Heitler, W., 1936 a, *Quantum Theory of Radiation* (Oxford), 119; 1936 b, *Ibid.*, 122 ; 1936 c, *Ibid.*, 115.
Hylleraas, E. A., 1937, *Zeits. f. Physik*, **106,** 395.
Johnston, J. E., 1939, *Proc. Camb. Phil. Soc.*, **35,** 108.
Jones, H., and Mott, N. F., 1937, *Proc. Roy. Soc.* A, **162,** 49.
Kingston, R. H., 1951, *Phys. Rev.*, **84,** 944.
Laue, M. V., 1914, *Ann. d. Phys.*, **44,** 1197.
Meier, W., 1910, *Ann. d. Phys.*, **31,** 1017.
Mott, N. F., and Jones, H., 1936, *Metals and Alloys* (Oxford).
Mott, N. F., and Massey, H. S. W., 1949 a, *Atomic Collisions* (Oxford), 111 ; 1949 b, *Ibid.*, 31.
Mott, N. F., and Sneddon, I. N., 1948, *Wave Mechanics* (Oxford), p. 264.
O'Bryan, H. M., 1940, *Phys. Rev.*, **57,** 995.
Pauling, L., and Wilson, E. B., 1935, *Quantum Mechanics* (New York), p. 240.
Skinner, H. W. B., 1940, *Phil. Trans. Roy. Soc.*, **239,** 95.
Skinner, H. W. B., and Johnston, J. E., 1937, *Proc. Roy. Soc.*, **161,** 420.
Slater, J. C., 1930, *Phys. Rev.*, **37,** 57 ; 1951, **81,** 385.
Stobbe, M., 1930, *Ann. d. Phys.*, **7,** 661.
Veldkamp, J., 1935, *Physica*, **2,** 25.
Whittaker, E. T., and Watson, G. N., 1950, *Modern Analysis* (Cambridge), p. 172.

CXV. *An Interferometric Investigation of the Abrasion Hardness Properties of Diamond*

By Eileen M. Wilks*
Royal Holloway College, University of London†

[Received August 11, 1952]

Summary

The abrasion properties of several diamonds have been investigated using a multiple beam interferometric technique. The cube, dodecahedron and octahedron planes were abraded in various directions using a special micro abrasion tester. Abrasion only occurred in directions parallel to crystallographic axes ; that is in four directions on cube planes and in two on dodecahedron planes, whereas no abrasion at all was produced on either natural or cleavage octahedron planes. Not all these directions were equally favourable ; although it was found that an exact orientation reduced the vector effect. Taking the most favourable direction in each case the dodecahedron surfaces were easier to abrade than the cube surfaces.

In addition the interferograms show that the abrading wheel moved in both lateral and longitudinal directions whilst abrading and at the same time probably suffered elastic deformation. It can be seen that the sides of an abrasion were highly polished and that there was no large scale deformation of the surrounding surface.

§1. Introduction

Diamond polishers have known for several centuries that only certain faces of diamond can be polished, and that not all directions on these faces are equally favourable for grinding. However, no systematic study was undertaken until 1920 when Tolkowsky investigated the rate of grinding in different directions, and compared the rate of grinding for different crystallographic faces. Further investigations and interpretations have been advanced by Stott (1931), Kraus and Slawson (1939), Whittaker and Slawson (1946), Slawson and Kohn (1951). These workers concluded that polishing occurs most easily in directions parallel to crystallographic axes. Moreover, on a given crystallographic face it is equally easy to produce abrasion in each of these directions. For example, a cube face, parallel to two crystallographic axes, has four equally favourable directions for grinding. They found that the optimum grinding directions on a dodecahedron face were easier to abrade than the corresponding directions on a cube face : there was also a gradual change in hardness in moving from one crystallographic face to another, except

* Communicated by the Author.
† Now at the Clarendon Laboratory, University of Oxford.

that the octahedron face, equally inclined to all three crystallographic axes, could not be abraded in any direction. However, according to some diamond polishers (Grodzinski and Stern 1949 a, Slawson 1946) there is a pronounced vector effect of abrasion hardness; on a cube face for example, abrasion will only occur in one direction.

The above results were obtained from the polishing of entire crystallographic faces; the weight of diamond removed in a given time being determined. The disadvantage of this method is that any variations in abrasion hardness due to the presence of submicroscopic twins (Sutton 1928) will not be revealed. By using a micro-abrasion tester (Grodzinski and Stern 1949 b) for the present investigation, only a small part of a surface was abraded at a time, and since the length of the abrasion made was only of the order of 10^{-2} cm several abrasions could be made on each face (the average area of a face being 4 sq. mm). Further, any submicroscopic twins would be revealed and it would therefore be possible to decide whether there were in fact fundamental variations in hardness properties between diamonds from different sources.

§ 2. Experimental Technique

The Grodzinski micro-abrasion tester consists primarily of a double conical grinding wheel of 1 inch diameter with included angle of 110°. It was charged with diamond powder of 0–2 micron size, and rotated at approximately 10 000 r.p.m. The diamond to be tested was so mounted on a balance arm that it was held against the abrading wheel by a force of 75 g wt. The time of each test run was 15 sec. The abrasions were made by W. Stern of the Diamond Research Department of the Industrial Distributors (Sales) Limited, and the abraded surfaces were studied using multiple beam interference techniques (Tolansky 1948). The depth of each abrasion was determined from the number of Fizeau type fringes contained within each outline, but when the depths were less than a light wave, fringes of equal chromatic order were used (Tolansky 1948). Examples of the two forms of fringes are shown in figs. 1 and 2 (figs. 1–3, Plate LXXII). Figure 1 (magnification $\times 44$) shows Fizeau type fringes on the cube face of a glassy diamond, with arrows indicating the direction of abrasion, and fig. 2 ($\times 220$) shows the fringes of equal chromatic order across the most shallow abrasion on this face. The length and width of each abrasion were determined from photomicrographs, but for reasons to be discussed later, the depth measurement was considered to give the more accurate estimate of the abrasion hardness; the latter being finally expressed as the depth per revolution of the abrading wheel.

Two series of tests were made: in the first, two diamonds were tested, a glassy and an alluvial, both originating from South Africa. Cube and dodecahedron faces of both stones were abraded, several directions on each face being tested. The natural and cleavage octahedron faces of the glassy were also tested. These faces were first orientated visually from an inspection of the natural features, but in a second series the surfaces were

orientated by x-ray diffraction. In this series cube faces of four diamonds were abraded, including the glassy and alluvial stones of the first series, together with two stones from the Belgian Congo (ref. No. BCK No. 27a and BCK No. 71a).

The hardness of the diamond had a most noticeable effect on the cast-iron wheel, which became very blunted with use. Also the amount of diamond powder on the wheel decreased with use, but although the wheel was recharged with powder after an average of 20 test runs, the wheel was not reground until it had become excessively blunt. This had a progressive effect on succeeding abrasions, for example, on the cube face of the diamond BCK No. 71a, four runs were made in the same direction, and the depth decreased steadily with successive runs. However, the variations in abrasion hardness between different faces, and between different directions on the same face, were always greater than the variation due to this lack of uniformity in the abrading conditions. Therefore, although no strictly quantitative assessment can be drawn from the experiments, a good qualitative estimate of abrasion hardness variations can be obtained.

§3. Experimental Results

With respect to the cube and dodecahedron faces of the glassy and alluvial stones, it was found that abrasion occurred *only* when the direction of grinding was parallel to a crystallographic axis, and when the directions of least resistance to abrasion were compared, it was found that the dodecahedron faces were some 25% easier to abrade than the cube faces. These results are in qualitative accord with the results of Tolkowsky and Slawson (*loc. cit.*). However, an interesting feature emerged when the abrasion hardness values for the most favourable directions on the two stones were compared. The dodecahedron face of the glassy was easier to abrade than that of the alluvial yet on the other hand, the cube face of the glassy was more resistant than that of the alluvial. This fact makes suspect any general statement about the variation of abrasion hardness between different stones unless similar crystallographic faces are being compared. No abrasion mark was produced at all on either the natural or the cleavage octahedron faces of the glassy stone, in agreement with the traditional view that octahedron faces cannot be polished in any direction.

With respect to the cube faces, it was found that there was a marked difference in the apparent abrasion hardness values between the faces that had been set by visual examination, and those that had been set by x-ray diffraction methods, the effect being more pronounced on the alluvial stone. Table 1 below gives the average hardness values expressed as depth of abrasion per revolution for the alluvial stone.

It is seen that although abrasion occurred in all directions parallel to the two crystallographic axes, on the visually orientated face there is one markedly favourable direction of grinding together with a pronounced vector effect, whereas on the face orientated by x-ray diffraction there are two equally favourable directions of grinding and the vector effect has

been reduced. These results may partly resolve the discrepancy between theoretical prediction and familiar experiment, for it would appear that diamond polishers may seldom be abrading the true cube faces. The abrasions produced on the glassy and on the two Belgian Congo stones which were orientated by x-ray diffraction showed similar results, in that there were two directions at right angles to each other along which the abrasion hardness values were nearly equal, whereas in directions at 180° to them the resistance to abrasion was high.

Table 1. Abrasion Hardness Expressed as Depth rev^{-1}

Alluvial stone	Abrasion direction	Visually orientated $cm \times 10^7$	Orientated by x-ray diffraction $cm \times 10^7$
(1)	Parallel to a crystallographic axis	0·37	0·75
(2)	90° to (1)	0·97	0·75
(3)	180° to (2)	1·13	1·33
(4)	180° to (1)	1·57	1·34

Table 2 shows the average abrasion hardness values in the most favourable directions for the cube faces of the four different stones.

Table 2

Diamond	Hardness value
BCK No. 27a	0·80
Glassy	1·13
Alluvial	1·34
BCK No. 71a	1·39

(The hardness value is expressed as the depth of abrasion per revolution, the unit being $cm \times 10^{-7}$ rev^{-1}.)

The magnitude of the differences between these values, as much as 40% between the two stones from the Belgian Congo, indicate a fundamental variation in the hardness properties of diamond as distinct from variations due to the presence of sub-microscopic twins.

The dodecahedron faces of the glassy and alluvial stones were orientated visually, and fig. 3 (×36) shows the Fizeau type fringes over the surface of the alluvial stone, with the arrows indicating the direction of abrasion. It was found that there was a distinct difference between the results obtained from the two stones. On the alluvial stone, abrasion occurred in both directions parallel to the crystallographic axis whereas on the glassy it

only occurred along one of these directions, although a small abrasion was effected at right angles to the axis. Moreover this favourable direction on the glassy was so easy to abrade that the time of abrasion had to be reduced to 5 sec. However, these differences may have been caused by an inexact orientation of the two faces.

In addition to the hardness determinations, the interferograms of the abrasions yielded some information concerning the effect on the abrading wheel of the hard diamond surface, for the geometrical outlines of the abrasions revealed several anomalies. Firstly, the included angle of the minor axis was found to be 176°, the angle being approximately the same for all the abrasions, yet the included angle of the double conical wheel measured before and after the test runs, was only 110°. Then, when the depths of the abrasions were related to the corresponding lengths and widths, it was found that although there was a general correlation between the length and depth, there was no such correlation between the width and the depth, occasionally the widest abrasion was also the shallowest. Finally when the measured depth was compared with the depth computed from the length of the abrasion and the geometry of the cast iron wheel, it was found that the experimental depths were always less than the computed depths, by between 10 and 25% ; the greater factor applying to the shorter abrasions.

Three causes were considered for these anomalies,

(*a*) that elastic deformation occurred during the abrading test,

(*b*) that the wheel moved laterally in its bearings during the test, and

(*c*) that the wheel moved longitudinally during the test.

Subsidiary experiments were therefore carried out, using a wheel with a controlled lateral movement. The resulting width of the abrasion was considerably increased, and the form of the abrasion showed that the wheel had slid over the surface, making several slight abrasion marks, until a groove was formed ; once the wheel was in the groove, the lateral movement ceased. It was therefore thought that the excessive width of the abrasions was caused by a mechanical flexibility in the supports of the wheel which had resulted in a lateral movement over the surface of the diamond ; the effect being more pronounced the more resistant the surface.

To investigate the difference between the computed and experimental depths, abrasions were made on a glass surface, and it was found that the computed depth was twice as large as the experimental depth. Since the wheel could not have been elastically deformed while abrading the glass, it was concluded that there was a longitudinal motion of the wheel which resulted in an increased abrasion length. It is probable that the uneven outline at one end of the abrasions which is shown by the interferograms is due to this longitudinal motion of the wheel.

Thus the mechanical movements of the wheel could have caused the dimensional anomalies of the abrasions ; the longitudinal motion affecting the length, and the lateral motion affecting the width ; the latter motion also tending to increase the included angle of the minor axis. Yet, owing

to the large discrepancy between the included angle of the cast iron wheel and that of the minor axis of the abrasion, the possibility of elastic deformation of the wheel must not be completely disregarded.

The interferograms of the abrasions also enable some conclusions to be drawn concerning the mechanism of abrasion. The figures show that the fringes within all the abrasions were extremely sharp and smooth, indicating that the surfaces within the abrasions were all highly polished, even though there is an anisotropy in the hardness property. When cube and dodecahedron faces were ground in directions at 45° and 90° respectively to the axis, the resulting surfaces were so uneven that no fringes were obtained, thus confirming earlier suggestions (Finch 1937, Bowden and Tabor 1950) that diamond is ground solely by the removal of its constituent particles.

Acknowledgments

My thanks are due to Messrs. P. Grodzinski and W. Stern of the Diamond Research Department of Industrial Distributors (Sales) Limited for many helpful discussions, and to Professor Tolansky for his constant interest in this work. Part of this work was done while holding a grant from the Ministry of Education.

References

Bowden and Tabor, 1950, *The Friction and Lubrication of Solids* (Oxford).
Finch, 1937, *Gemmologist*, **7**, 82, 192.
Grodzinski and Stern, 1949 a, *Ind. Diam. Rev.*, **9**, 122; 1949 b, *Nature, Lond.*, **164**, 193.
Kraus and Slawson, 1939, *Amer. Min.*, **24**, 661.
Slawson, 1946, *Amer. Min.*, **31**, 163.
Slawson and Kohn, 1950, *Ind. Diam. Rev.*, **10**, 168.
Stott, 1931, *N.P.L. Collected Researches*, **24**, 1.
Sutton, 1928, *The Diamond* (London).
Tolansky, 1948, *Multiple Beam Interferometry of Surfaces and Films* (Oxford).
Tolkowsky, 1920, *D.Sc. Thesis*, London.
Whittaker and Slawson, 1946, *Amer. Min.*, **31**, 143.

CXVI. *The Beta-Gamma Angular Correlation of ^{170}Thulium*

By H. Rose
Cavendish Laboratory, Cambridge*

[Received September 8, 1952]

Abstract

A thin magnetic lens beta-ray spectrometer has been modified to enable the beta-gamma angular correlation of ^{170}Tm to be measured as a function of beta-ray energy. Assuming that a single matrix element is operative in the low energy beta-ray transition of ^{170}Tm, the results give a good agreement with the theoretical predictions for the matrix element $\int\boldsymbol{\beta\sigma}\times\mathbf{r}$ occurring in the first forbidden tensor interaction but appear to rule out the corresponding $\int\boldsymbol{\sigma}\times\mathbf{r}$ of the first forbidden axial vector interaction. It is, however, possible that a suitable mixture of matrix elements might also fit the experimental data.

§1. Introduction

A beta-gamma angular correlation in the decay of ^{170}Tm has been discovered by Novey (1950). Assuming the correlation function $W(\theta)=1+a\cos^2\theta$, he obtained for a the value $-0{\cdot}259\pm0{\cdot}053$. However, as he points out, this result is somewhat uncertain on an absolute basis because his experiment was not carried out in a vacuum, and scattering in air would be expected to reduce the correlation.

The existence of the correlation implies that the low energy beta-ray spectrum of ^{170}Tm (see fig. 1) cannot be an allowed transition, and a recent investigation into the shape of this spectrum by Richmond and Rose (1952) has indicated that the transition is first forbidden with axial vector or tensor interaction, the latter interaction being slightly favoured. By investigating the beta-gamma angular correlation as a function of beta energy it has been possible to draw more definite conclusions regarding the type of interaction obeyed in the low energy beta-ray transition.

§2. Experimental Details

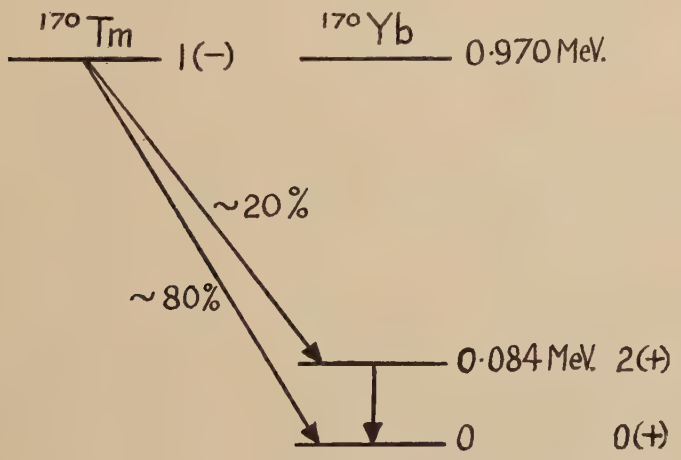
A thin magnetic lens beta-ray spectrometer was used to focus beta-particles of specific energies while gamma-rays from the same source were detected with a NaI crystal which could be rotated about the source. The source itself was mounted at the centre of a source-holder which was essentially a hemisphere of aluminium, so that the gamma-rays penetrated equal thicknesses of aluminium at all angles of detection.

* Communicated by Mr. E. S. Shire.

The spectrometer was set at a momentum resolution of 9% to give a relatively large transmission. At this setting the spectrometer baffles limited the detection of beta-particles to those leaving the source at angles between 3° and 8° to the spectrometer axis.

The beta-particles were counted by an anthracene crystal which was mounted on the vacuum side of a glass disc sealing the exit hole in the end-plate, and the light pulses were piped off by a Perspex rod 18 cm long to a photomultiplier (E.M.I. type 6262). The thallium-activated NaI crystal which detected the gamma-rays was about 4·5 cm^2 in cross-section and 1 cm thick and subtended a solid angle of 0·08 steradian at the source. The crystal was sheathed in a lead cylinder 1·5 cm thick to screen it from all radiation other than that coming directly from the source. A 25 cm Perspex rod acted as a light-guide between the NaI crystal and a second photomultiplier, the whole being capable of rotation through 90° on either side of the spectrometer axis.

Fig. 1



Decay scheme of ^{170}Tm.

The Perspex rods acted not only as light-guides but served also to remove the multipliers to some distance from the spectrometer lens coil. It was then possible to provide sufficient magnetic shielding to reduce the effect of the magnetic field of the coil on the multipliers to an inappreciable amount over the range of spectrometer currents used in the experiment without resorting to steel end-plates and a steel tank enclosure for the coil as was necessary in the beta-gamma angular correlation investigations of Stevenson and Deutsch (1951).

Coincidences were recorded between focused beta-particles and gamma-rays with the gamma detector at 90°, 180° or 270° to the spectrometer axis. To minimize errors due to changing efficiencies the angle of the gamma counter was changed and the observations repeated a considerable number of times for each beta-ray energy selected. The coincidence rates

were normalized with respect to the single counting rates in the gamma counter. The resolving time of the coincidence circuit used was 8×10^{-8} sec.

§3. Experimental Results

To ensure that the experimental arrangement did not suggest a correlation where none existed, tests were first carried out on ^{134}Cs and ^{198}Au, both of which are known to have isotropic correlations. (Stevenson and Deutsch 1951, Frankel 1950).

^{134}Cs was deposited as an aqueous solution on to 0·25 mg cm^{-2} aluminium foil and evaporated down to form a source of thickness about 0·5 mg cm^{-2}. The ^{198}Au source was electro-deposited on to a similar source backing and was only about 30 μg cm^{-2} thick. Both sources covered about 1 cm^2 in area.

No correlation was observed in the decay of either source. The results obtained are given in table 1.

Table 1

Isotope	Energy of beta-rays (MeV)	a (assuming $W(\theta)=1+a\cos^2\theta$)
^{134}Cs	0·300	$-0{\cdot}006\pm0{\cdot}024$
,,	0·489	$+0{\cdot}021\pm0{\cdot}032$
^{198}Au	0·614	$+0{\cdot}013\pm0{\cdot}035$

From some 'Specpure' Tm_2O_3 of high specific activity (4 mc mg^{-1}) a source of thickness 1 mg cm^{-2} and area 1 cm^2 was prepared on 0·25 mg cm^{-2} aluminium foil. A 1 g cm^{-2} brass absorber was placed in front of the NaI crystal. This transmitted about half of the gamma-rays but absorbed almost all the K x-radiation of ^{170}Yb, which was expected to be isotropic with respect to the beta-rays (Novey 1950).

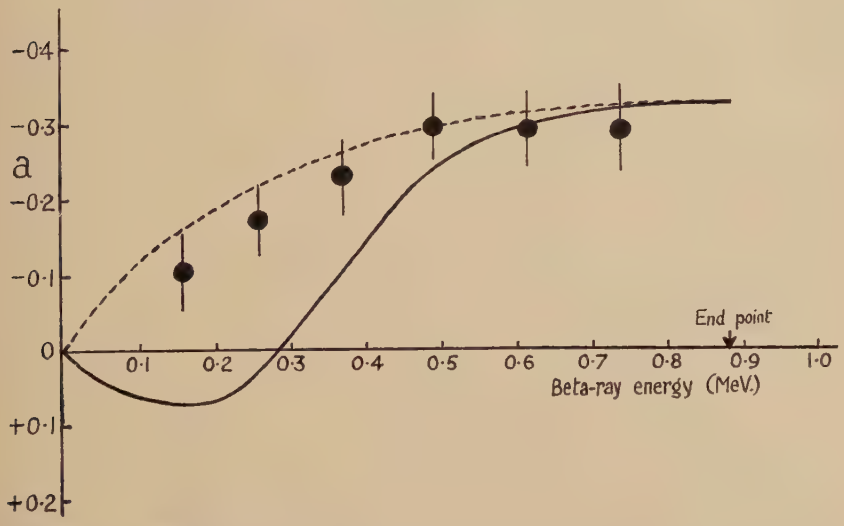
The 84 kev gamma-rays are to a large extent internally converted, leaving only a few per cent for detection. Because of this relatively few genuine coincidences were observed, and in order to obtain the statistics shown in fig. 2 it was necessary to count for periods of 40 hours or more at each point on the beta-ray spectrum.

The angular correlation function $W(\theta)$ is of the form $1+a\cos^2\theta$ for all first forbidden beta transitions followed by a gamma-ray (Falkoff and Uhlenbeck 1950). The values obtained for the differential correlation coefficient a at various points on the beta-ray spectrum, after correction both for the finite solid angles of acceptance of the detectors and for the spectrometer resolution, are given in fig. 2. The errors indicated are the probable errors based on the number of counts.

Assuming an allowed spectrum shape (Richmond and Rose 1952) for the low energy beta transition an integral correlation coefficient

$a=-0{\cdot}19\pm0{\cdot}04$ was calculated on the basis of the experimentally observed differential coefficients of fig. 2. This appears at first glance to be rather lower than Novey's reported value, but it must be borne in mind that in his experiment he used a 3 mg cm^{-2} mica window over the beta detector to prevent counting of conversion electrons. Consequently his value is not a true integral correlation coefficient but represents a correlation integrated over beta energies between about 150 kev and the end point. Such a procedure, when applied to the results of fig. 2, yields a value for a of $-0{\cdot}24\pm0{\cdot}04$, which is in good agreement with that of Novey.

Fig. 2



Beta-gamma angular correlation of ^{170}Tm. The solid line denotes the theoretical correlation ($Z=0$) for the first forbidden matrix element $\int\boldsymbol{\sigma}\times\boldsymbol{r}$ occurring in the axial vector interaction, the dotted line that for the first forbidden matrix element $\int\beta\boldsymbol{\sigma}\times\boldsymbol{r}$ of the tensor interaction.

§4. Discussion

Since ^{170}Yb is an even–even nucleus its ground state should have spin zero and even parity. Measurements of conversion electron intensities (Grant and Richmond 1949) and of the lifetime of the 84 kev excited level (Bell and Graham 1950) have shown that the 84 kev transition is electric quadrupole and, therefore, the excited level has spin 2 and even parity. The shape of the low energy beta-ray spectrum (Richmond and Rose 1952) indicates that the transition is first forbidden and requires ^{170}Tm to have spin 1 and odd parity.

With this assignment of angular momenta the only first forbidden nuclear matrix elements which can give the negative coefficient a are $\int\boldsymbol{\sigma}\times\boldsymbol{r}$ and $\int\beta\boldsymbol{\sigma}\times\boldsymbol{r}$ occurring in the axial vector and tensor interactions respectively. The theoretical correlations (Falkoff and Uhlenbeck 1950) to be expected for these two matrix elements are plotted in fig. 2. It

can be seen that whereas they yield the same correlation for the high energy beta-particles, they differ considerably at intermediate energies. The experimental points agree well with the theory for the tensor interaction and seem to eliminate the axial vector as a possibility.

The theory of Falkoff and Uhlenbeck (1950) is strictly correct only for the case $Z=0$, and little is known about the precise influence of Z. Although the effect of Z is generally quite small (Fuchs and Lennox 1950) it can substantially lower the correlation in absolute value from that computed in the $Z=0$ approximation for cases where matrix elements like $\int \boldsymbol{\sigma} \times \mathbf{r}$ are involved (Stevenson and Deutsch 1951). This seems to rule out completely the possibility of the axial vector being the true interaction in the low energy beta transition of ^{170}Tm but might explain why the experimental points fall somewhat below the theoretical curve for the tensor interaction. There is also the possibility that a suitable mixture of first forbidden matrix elements might fit the experimental results, but no theoretical calculations of explicit correlation functions assuming such mixtures are at present available for comparison.

I wish to express my thanks to the Atomic Energy Research Establishment, Harwell, for providing the active sample of ' Specpure ' Tm_2O_3 used in this experiment, and also to Dr. A. G. Maddock for assistance in the preparation of the sources.

References

Bell, R. E., and Graham, R. L., 1950, *Phys. Rev.*, **78,** 490.
Falkoff, D. L., and Uhlenbeck, G. E., 1950, *Phys. Rev.*, **79,** 334.
Frankel, S., 1950, *Phys. Rev.*, **77,** 747.
Fuchs, M., and Lennox, E. S., 1950, *Phys. Rev.*, **79,** 221.
Grant, P. J., and Richmond, R., 1949, *Proc. Phys. Soc.* A, **62,** 573.
Novey, T. B., 1950, *Phys. Rev.*, **78,** 66.
Richmond, R., and Rose, H., 1952, *Phil. Mag.*, **43,** 367.
Stevenson, D. T., and Deutsch, M., 1951, *Phys. Rev.*, **83,** 1202.

CXVII. *A Theory of Work-hardening of Metal Crystals*

By N. F. Mott, F.R.S.
H. H. Wills Physical Laboratory, University of Bristol*

[Received August 8, 1952]

Summary

A theory of work-hardening is put forward ; the salient points are as follows. There are two types of hardening ; the normal rapid hardening of cubic metals and the slower hardening characteristic of hexagonal crystals and cubic metals under conditions of easy glide. A quantitative theory of the rapid hardening is given. This does not occur unless dislocations are retained in the slip planes. The retained dislocations are supposed to occur in groups of about 1000 at the ends of the slip-bands. The strains round these groups are responsible for hardening as in Taylor's (1934) theory. The obstacles which stop slip are sessile dislocations, and the formation of sessile dislocations stabilizes the groups of dislocations. It is shown that the formation of slip lines must be a dynamic process, and on this assumption a parabolic stress–strain curve is deduced. Some discussion is given of the formation of vacancies by moving dislocations, and it is shown that this can account for the clustering of slip lines observed under the electron microscope. The theory agrees qualitatively with observations on the amount of retained energy in a cold-worked metal.

§1. Introduction

The aim of this paper is to give a theory of work-hardening of metal crystals, based on the properties of dislocations. The basic assumption made is that a crystal contains a network of dislocations,† and that a dislocation in an otherwise perfect crystal moves when the crystal is subjected to a small stress. We shall attempt from these assumptions to build up a quantitative theory for close-packed cubic crystals, and shall make a few qualitative remarks about hexagonal crystals.

The first theory of work-hardening based on dislocations is that due to Taylor (1934). Qualitatively this may be expressed as follows. The stress round an edge dislocation is of the form

$$\sigma = \pm (Gb/r) f(\theta), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

* Communicated by the Author.

† It is not suggested that it is inherently impossible to prepare crystals without dislocations ; the dislocations in real crystals are a result of the method of growth. Some of the factors involved have been summarized by Frank (1952). A recent note by Herring and Galt (1952) shows that very thin fibres formed on tin crystals have yield strengths approaching the values expected for a material without dislocations.

where G is the shear modulus, b the Burgers vector, r the distance from the dislocation line and $f(\theta)$ a function of angle. If, then, a cold-worked solid contains a random array of dislocations of both signs at a mean distance l from each other, there will be in the material a stress σ_i, varying with position in some roughly sinusoidal way, of which the magnitude is of order

$$\sigma_i = Gb/l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If each dislocation of the array is in some way ' stuck ', any additional dislocation introduced into the lattice cannot move through distances large compared with l unless the applied stress σ exceeds σ_i. σ_i is then identified with the flow stress of the crystal at any point of the stress–strain curve. It is supposed in Taylor's model that dislocations are constantly being created while the substance is being deformed, and that they move a distance L before they are stuck. The plastic strain ϵ is thus

$$\epsilon = Lb/l^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Eliminating l between (2) and (3), we find

$$\frac{\sigma_i}{G} = \sqrt{\frac{b}{L}}\epsilon^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

To obtain agreement with experiment for the stress–strain curve of aluminium, Taylor had to assume that L is about 10^{-4} cm, and he deduced that, after a strain of order unity, $l \sim 3 \times 10^{-6}$ cm. The concept of an oscillating internal field σ_i has been adapted by Mott and Nabarro (1948) and Nabarro (1952) to the explanation of age-hardening and the high yield-stress of alloys ; while an explanation of yield points in iron and similar materials has been given by Cottrell (1948 ; see also Cottrell and Bilby 1949) in terms of the segregation of impurities actually in or very near the dislocations.

The explanation of strain hardening which we shall give is an extension of Taylor's ; we suppose that the hardness is due to an internal field due to dislocations generated in the process of plastic flow. The actual distribution, however, is very different from that of Taylor, dislocations of like sign being grouped together, some hundreds being in each group. It shares with Taylor's theory, however, the feature that the hardening is, to a first approximation,* independent of the direction in which slip has occurred ; in a cubic crystal dislocations generated by slip along one set of (111) planes harden the crystal for slip along another set of (111) planes.

To complete Taylor's theory, we have to ask :

(*a*) How are the dislocations generated during slip ? For this a satisfactory mechanism has been proposed by Frank and Read (1950), and will be discussed in the next section.

* This point has been examined further by Stroh (1952) who in a paper to be published shortly finds a small dependence.

(*b*) How do the dislocations become 'stuck' after moving a certain distance, what limits slip on a given plane, and why do not the dislocations move back when the stress is reversed? Here we propose different mechanisms for the edge and screw components of a dislocation loop. The edge components in a cubic close-packed crystal become stuck by combining with edge dislocations on other octahedral planes, as proposed by Lomer (1951) and Cottrell (1952). A join of two such dislocations produces what we call a sessile dislocation. This cannot happen in hexagonal crystals, which is why they harden so much more slowly. The screw dislocations are probably stopped by a process connected with the generation of vacancies by moving dislocations (Seitz 1952). These vacancies also play a large role in the temperature-dependence of strain hardening and in the increase in electrical resistance due to strain hardening.

Some of the experimental facts which have to be explained by any theory are:

(*a*) The occurrence of slip lamellae.

(*b*) The fact that at low temperatures, in aluminium at any rate, a slip lamella consists of slip *on one plane* through distances of the order 1200–1800 Å. As the temperature, however, is raised, the visible slip bands are found to consist of clusters of these elementary steps, about 200 Å apart and with roughly the same slip on each (Heidenreich and Shockley 1948, Brown 1949, 1951, 1952).

(*c*) The very considerable temperature-dependence of hardening.

(*d*) The form of the stress–strain curve.

(*e*) The formation of deformation bands and asterisms in cold-worked cubic metals, and their absence in hexagonal crystals.

(*f*) The phenomenon of easy glide (Andrade and Henderson 1951, Lücke and Lange 1952).

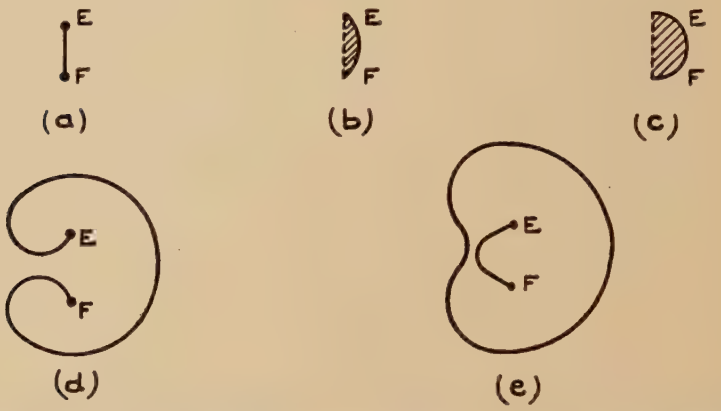
§ 2. The Frank–Read Source and the Origin of Slip Bands

Several accounts have already appeared of the explanation given by Frank and Read of the origin of slip bands (Frank and Read 1950, Frank 1951, Mott 1951). The model is illustrated in fig. 1. It is based on the concept of a dislocation line **EF** lying in a slip plane, which at the points **E**, **F** moves on to other planes lying at an angle to it. We consider movement in which **EF**, rather than its continuation in the other planes, is displaced.

Consider then what will happen to the line **EF**, when the material is stressed in such a way as to cause slip. If the line moves from the position shown in fig. 1 (*a*) to that shown in fig. 1 (*b*), the material over the area shown shaded in fig. 1 (*b*) slips relative to that underneath it The line will move in such a way as to extend this area. On the other hand the points **E**, **F** are locked. Successive forms of the line are shown in figs. 1 (*a*) to (*e*). It will be seen that ultimately a complete loop is generated, and that at the same time the line **EF** is formed again.

A line such as EF is thus capable of generating an infinite series of loops, when a large enough stress is applied. Moreover, it will be shown later in this section that the stress is of the order of that for which yield of actual metal crystals occurs.

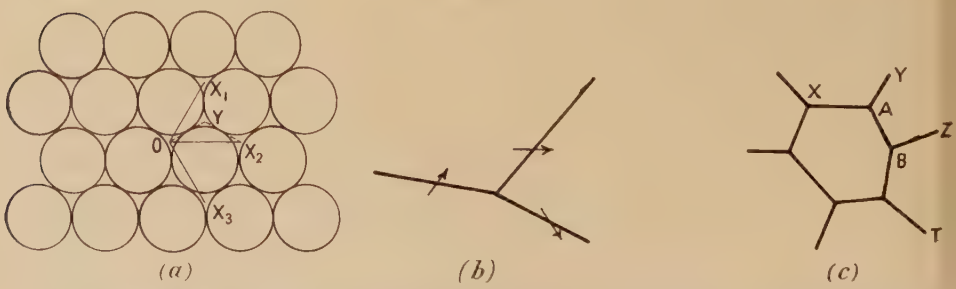
Fig. 1



Successive stages in the formation of a dislocation ring from a Frank–Read source.

For close packed cubic crystals—to which much of this article will be devoted—the model becomes more plausible when we consider the Burgers vectors possible for slip in the (111) planes. Such a plane is shown in fig. 2 (*a*). The possible values of the Burgers vector are OX_1, OX_2 and OX_3. It is clear that the displacement when two dislocations with vectors OX_1, OX_3 move across the plane is the same as if a dislocation with vector OX_2 moves across it. From this it follows that it is geometrically possible for three dislocations to meet in a point, as in fig. 2 (*b*), the Burgers vectors of the three being respectively OX_1, OX_2, OX_3. Further, it is obviously not necessary that the three lines should all lie in the same slip plane.

Fig. 2



(*a*) A close-packed plane, showing the Burgers vector for complete and half dislocations. (*b*) Join of three dislocations. (*c*) Network of dislocations (from Mott 1951).

A possible arrangement of dislocations in a cubic close-packed solid, and probably the most stable in view of the line energy of dislocations, is therefore a network such as that shown in fig. 2 (*c*). This is the model that we shall adopt throughout this report for the arrangement of dislocations in an annealed metal of close-packed structure; it will probably replace the old model of a 'mosaic crystal' so far as the interpretation of intensity measurements in the reflection of x-rays is concerned. With this model, any element such as **AB** can act as a Frank–Read source.

There are certain further points that we may note about this model. These are:

Extended Dislocations. As first pointed out by Heidenreich and Shockley (1948), a dislocation in a close-packed plane with Burgers vector say OX_2 in fig. 2 (*a*), can and will split into 'half-dislocations' with Burgers vectors OY, YX_2. These two half dislocations repel each other; it is easily seen that the elastic energy, proportional to the square of the Burgers vector, is diminished when they separate in the ratio

$$2OY^2/OX_2^2$$

and hence in the ratio $\frac{2}{3}:1$. On the other hand, when they separate they create a stacking fault between them, the energy of which increases with the distance between them. According to Heidenreich and Shockley, they will be in equilibrium when separated by about 2×10^{-7} cm.

As a result of this dissociation, the dislocations in close-packed planes will have low energy, and the elements such as **AB** (fig. 2 (*c*)) in cubic close-packed structures will mainly be in close-packed planes. This is believed to be the reason why glide in close-packed structures shows a marked preference for close-packed planes, while for materials with some other structures (e.g. body-centred metals, silver halides), various planes show glide.

Yield Points. In a pure single crystal flow will begin when the applied stress is great enough to initiate the process illustrated in fig. 1, namely the generation of loops at the most favourably situated Frank–Read source. The force resisting the generation of loops is due to the energy per unit length, or tension of a dislocation. If σ is the resolved part of the stress about a line lying in the slip plane parallel to the Burgers vector, the force exerted on unit length of the dislocation is σb. If W is the energy of the line per unit length, the resultant force in the opposite direction is W/ρ, where ρ is the radius of curvature. The maximum value which $1/\rho$ can reach as the loop expands is $1/2l$, where l is the length **EF** of the source. Thus the source will generate loops if

$$\sigma \geqslant W/2lb.$$

Putting in estimated numerical values for W for isotropic materials, Frank (1950) finds that in practical cases $W \sim 2Gb^2$, so that the condition for yield is

$$\sigma \geqslant Gb/l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

For the best single crystal of, for example, silver (Andrade and Henderson 1951), yield points σ of the order 50 g/mm^2 are observed, giving values of l of about 3×10^{-3} cm ; for aluminium Rosi and Mathewson (1951) observed yield points of the order 200 g/mm^2 give $l\sim3\times10^{-4}$ cm. We emphasize that the values of l deduced are for the largest values of the length AB in the dislocation network discussed ; the average value may be smaller. No theory has, as yet, been given which enables us to estimate the mean value of l that will result from a given method of preparing a crystal.

There are two further effects to be considered. Firstly, there will exist a random internal stress due to other dislocations or impurities which changes sign in a distance λ smaller than the estimated values of l ; this effectively introduces a frictional force, say $\sigma_0 b$ per unit length resisting the motion of a dislocation ; a dislocation initially at rest will not move unless the applied stress exceeds σ_0, and a Frank–Read source will not generate dislocations unless

$$\sigma>\sigma_0+Gb/l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Secondly we have to consider the locking effect of impurities which have by diffusion moved into the immediate neighbourhood of a dislocation. This effect and its relation to yield points have been considered in a number of papers by Cottrell and his co-workers (Cottrell 1948, Cottrell and Bilby 1949). For our purpose we need to emphasize the following effects of such locking :

(*a*) The stress (σ_{Cot}) required to initiate a slip band is now simply that required to make a dislocation break away from its impurities, provided that σ_{Cot} is greater than σ_0+Gb/l. Once the dislocation has broken away, however, the resistance to further slip is still given by (6).

(*b*) Since σ_{Cot} will be the same for all loops and is thus independent of l, the conditions are favourable for sharp yield points.

(*c*) σ_{Cot} is very dependent on temperature (Cottrell and Bilby 1949), while σ_0 is very little dependent on temperature (Mott and Nabarro 1948).

It is of some interest to consider the effect of a network such as that considered here on the apparent elastic modulus. If the dislocations are locked in the Cottrell sense, there should be no movement for stresses less than σ_{Cot}. Up to this stress the material should be perfectly elastic with the theoretical shear modulus. In the absence of Cottrell locking, however, a stress greater than σ_0 but less than σ_0+Gb/l will lead to a reversible movement of any dislocations present in the material. If the material is highly purified so that σ_0 is very small, this will lead to an apparent change in the elastic modulus. We shall now estimate the magnitude of this for the network illustrated in fig. 2 (*c*). If the stress applied is σ in the plane considered, the dislocation takes up the form of a circle of radius ρ given by

$$\sigma b=W/\rho.$$

Substituting for W we find

$$\rho=2bG/\sigma.$$

The area swept out by the dislocation (the shaded area in fig. 1 (*b*)) is thus $\rho^2(\theta-\frac{1}{2}\sin 2\theta)$, where $\sin\theta=\frac{1}{2}l/\rho$. For small θ this reduces to

$$l^3/12\rho=l^3\sigma/24bG.$$

If there are N Frank–Read sources per unit volume, it follows that the shear produced by a stress σ is

$$Nl^3\sigma/24G. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Since Nl^3 is probably slightly less than unity, it follows that the correction to the elastic modulus is of the order of at most a few per cent, unless σ approaches the stress Gb/l.

It follows that the stress–strain curve of a pure substance having a number of Frank–Read sources shows a (reversible) strain slightly in excess of that given by the ideal elastic modulus. Under alternating stresses there will be a slight dissipation of energy, of amount proportional to the temperature, due to a mechanism first put forward by Eshelby (1949). If, however, there is any frictional resistance to the motion of dislocations, one will obtain a hysteresis loop (cf. Nabarro 1951). Such a resistance will certainly exist in alloys for the reasons already stated.

Fig. 3

Glide as the result of stress.

We may now sum up the elementary consequences which follow from the assumption that the crystal contains a number of Frank–Read sources, or a network of the type shown in fig. 2 (*c*). If the crystal is submitted to a uniform shear stress, then as the stress is increased a value will be reached at which one of the sources, that for which l is largest, begins to generate loops. The process will then continue indefinitely; nothing introduced into the model so far will stop it. Glide as shown in fig. 3 is thus to be expected, until owing to the change of shape of the crystal the stress is no longer uniform, when other sources would be brought into play. If by some mechanism slip on the first plane was stopped, the source with the next largest value of l would be brought into play. The next task of a theory is to explain why slip ever stops on a given plane. In particular we have to explain the distance of *c.* 2000 Å shown by the work of Heidenreich and Shockley and of Brown to be characteristic of slip on each active slip band.

§3. The Piling-up of Dislocations against a Barrier

The model outlined in the last section provides for a limited number of sources, say $10^{10}-10^{12}$ per cubic cm in an annealed crystal; each of these sources is capable of producing an unlimited number of dislocation loops when a large enough stress is applied. If in the material there are no barriers to the movement of dislocations, these loops will spread to the boundaries of the crystal where they will disappear, causing glide as in fig. 3. We shall have, however, to consider what happens when the material contains some internal surface, which acts as a barrier against which dislocations may pile up. Such a barrier may be, for instance, a grain boundary; later in this article we shall consider the influence of trapped or 'sessile' dislocations and of deformation bands.

To fix our ideas, we shall consider a row of dislocations, either edge or screw, moving in a single slip plane, piled up against a barrier under the action of a stress σ, as illustrated in fig. 4. This important case has been treated in detail by Eshelby, Frank and Nabarro (1951). The following results are obtained:

(*a*) If there are n dislocations present, the force on the leading dislocation is n times that on a single dislocation, namely $n\sigma b$ per unit length.

Fig. 4

Pile-up of dislocations against a barrier.

(*b*) The length L between the first dislocation and the barrier is given by

$$L=knGb/\pi\sigma, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where k is unity for a screw, $1-\nu$ for an edge dislocation, G is the shear modulus and ν is Poisson's ratio.

(*c*) The distance between the two leading dislocations is

$$d=1{\cdot}84\, d_0/n,$$

and that between the last two of order $d_0 n^{1/3}$ where

$$d_0=Gb/2\sigma k.$$

(*d*) The shear stress on the glide plane at a distance x ahead of the leading dislocation is practically the same as that due to this dislocation by itself in the region where $x \ll d$. At distances large compared with d but small compared with L the stress is approximately

$$\sigma\{1+(L/x)^{1/2}\}.$$

The stress concentration is thus large, the dislocations producing the same stress concentration as does a freely slipping crack of the same length.

(*e*) The total slip produced by the sequence is the same as would be produced if a single dislocation of strength nb moved a distance $3L/4$.

Some numerical values will be of interest here. First of a well, estimate the magnitude of the distance d_0, which is equal to the equilibrium distance between two dislocations of the same sign, one of which is held up by a barrier and the other forced towards it by a stress σ. For a metal such as Cd or Al, we may take $G \sim 2\times10^{11}$ dynes/cm^2 ; for the yield point of a pure metal, $\sigma \sim 100$ g/mm^2 $\sim 10^7$ dynes/cm ; this gives $d_0 \sim 3\times10^3 b \sim 5\times10^{-5}$ cm. For a crystal that has been heavily deformed the flow stress will be from 10 to 100 times greater and the magnitude of d_0 is 10^{-5} or 10^{-6} cm. Any theory which seeks to ascribe work-hardening to the interaction between dislocations fairly uniformly distributed (e.g. Taylor 1934) must assume that the distance between them is of this order, and thus considerably less than the distance between observed slip bands. One has to assume, *either* that in a cold-worked metal the space between the slip bands is filled up with dislocations, *or* that in the slip bands there are clusters of piled-up dislocations, formed as described above. Each of these clusters will behave at large distances like a dislocation of large Burgers vector. In the latter sections of this report we shall develop particularly the second of these hypotheses, though we believe that there will be some formation of dislocations between the slip bands.

A second point of interest is the high stress concentration that exists near the limit of a row of piled up dislocations. Frank (1951) has suggested that, when this occurs at a grain boundary, the stress may be great enough to generate loops. This point we believe to be important, and it will be considered in further detail in § 6. Here we wish to emphasize that such a process is likely to be temperature-dependent ; temperature can assist the formation of a loop in a region of very high stress.

When the surfaces of aluminium crystals which have been deformed at low temperatures are examined with the electron microscope, *single* steps of the order of 1000–2000 Å in height are observed. At higher temperatures the steps are clustered ; a hypothesis to account for this clustering will be given in § 8. At low temperatures, however, this suggests that about 1000 dislocations are produced by each Frank–Read source. The edge dislocations of one sign producing the visible slip band have come out of the crystal ; those of the opposite sign, at the opposite side of the ring, must in general be piled up against some barrier. We shall now estimate the distance from the surface that this barrier must be. Let us assume for the moment, then, that when each ring expands the screw elements go to the boundary of the crystal or at any rate go a large distance compared with the edge components, but that the latter are stopped at a barrier after moving a distance L. We have to ask how big this distance L must be in order to account for a value of n of *c.*1000.

We shall take the case of aluminium single crystals subjected to a shear stress of 500 g/mm^2 or 5×10^7 dynes/cm^2. If $G\sim3\times10^{11}$, and $n\sim1000$, then formula (8) gives $L=0{\cdot}05$ cm. The distance of the first barrier from the surface may be as much as twice this. Thus, for slip of 2000 Å to take place for the lowest stresses at which yield occurs, a slip distance L of nearly 1 mm is necessary.

We shall now discuss the *dynamics* of the formation of a slip line. From the theoretical point of view, it is not yet certain how much a moving dislocation in a perfect crystal is damped by the thermal vibration of the lattice (Leibfried 1949, 1950, Nabarro 1951). We believe, however, that the facts about the formation of slip bands show conclusively that dislocations must acquire a velocity not less than about half the speed of sound. For all the observations of slip bands (Brown 1951) show that the *elementary* slip band through *c.* 2000 Å is formed completely or not at all, though the formation of other members of a *cluster* may occur after some time. The evidence suggests, then, that the elementary slip band is formed quickly, and not gradually as the deformation proceeds. While a crystal is being strained, the stress rising slowly, slip from a given source begins then presumably when σ reaches the value of Gb/l for that source, and the band is formed before σ has risen appreciably above this value. If the dislocations moved slowly, say with one tenth of the velocity of sound, this would be impossible ; as soon as, say, Δn dislocations were piled up at the end of the slip distance L, they would exert a stress $\Delta nG/2\pi L$ at the source, *preventing* the generation of new dislocations. This would stop the source from acting long before the final value n was reached. If the motion of disclocations is heavily damped, we should expect n to increase gradually with σ ; it could not reach its full value until σ reached about twice the yield stress.

If on the other hand dislocations can move with nearly the speed of sound, once a Frank–Read source is set into action, it will go on generating dislocations under its own momentum, the momentum generated in moving from the configuration of fig. 1 (*e*) carrying the line **EF** to the configuration of fig. 1 (*c*). Generation of loops will then go on until the stress of the piled-up dislocation stops it, the whole process being over in a few microseconds.*

Moreover, if the process is rapid as described here, we can, assuming the presence of a barrier at a distance L from the source, account for the fact that the dislocations do not move back when the stress is taken off. This can be explained in terms of the formation of *sessile* dislocations, which is likely to occur as soon as a piled-up group of more than about

---

* A somewhat similar suggestion has been put forward by J. C. Fisher, E. W. Hart and R. H. Pry (to be published). These authors consider that the stress due to the still moving dislocations generated by the source will stop the source from acting after some hundred are formed, without their coming up against any barrier. I am indebted to Dr. Fisher for allowing me to see his manuscript prior to publication.

six dislocations is brought to rest, but will not occur while they are in motion. We have already stated that when two dislocations on two octahedral planes join, they form a sessile dislocation. A piled up group of n dislocations has a stress field round it of order $nGb/2\pi r$. There are other sources at a mean distance from it of order l, and a stress of order Gb/l is required to make any one of them operate. Thus a piled-up group of six is sufficient to cause plastic flow round it, and to produce one or more edge dislocations on a different plane which will form a sessile element with one of the six. We suppose, then, that this is what happens, and a proportion of about one in six of the dislocations piled up at the end of a slip plane become sessile, thus preventing any appreciable recovery of strain when the stress is removed.

We emphasize that this phenomenon cannot occur for hexagonal crystals, where there is only one set of slip planes.

It is clear that in ordinary plastic flow, only a small proportion of the available sources are used. For supposing that the slip distance for edge and screw dislocations is L, then the strain ϵ produced by the operation of N sources per cm^3 is

$$\epsilon = NLL'bn.$$

With $L \sim L' \sim 10^{-2}$ cm, $n \sim 1000$ and $\epsilon \sim 1$, this gives $N \sim 10^8$, considerably less than the value $10^{10}-10^{12}$ which we believe to represent the total number.

Evidence in favour of this conclusion is provided by the work of Brown and Honeycombe (1951), who have shown that the slip markings distant *c.*1 micron from each other are characteristic only of surfaces which have been mechanically polished, and that they do not appear on surfaces that have been electrolytically polished and etched. We believe, however, that the former markings are characteristic of the interior, and for the following reason. As first pointed out by Holloman (1952), dislocations which terminate on the surface of an otherwise perfect crystal will behave as though locked between the real locking point and its reflection in the surface. They will have, therefore, much longer effective values of l. If

$$l \sim 10^{-4} \text{ cm},$$

there will be per unit area of surface *c.* 10^8 sources which terminate in the way shown ; and if a stress is applied which will operate any of the sources in the interior, it will operate also most of the surface ones. If the dislocations generated move a distance 10^{-2} cm, we see that a density of slip lines 10^6 per cm is to be expected on the surface. Under such conditions it is natural that each line can only have a small displacement.

If the surface has been cold-worked or hardened in any other way, the situation is quite different ; sources on the surface do not move more easily than those on the interior, and if the hardened layer is thin compared with the slip distance (say 10^{-2} cm), the markings shown should be characteristic of the interior.

§ 4. Strain-hardening of Cubic Close-packed Crystals

We give in this section a theory, allied to that of G. I. Taylor, of the work-hardening of cubic close-packed crystals at low temperatures. The theory will be limited to one rather rapid hardening characteristic of impure crystals; the 'easy glide' sometimes obtained with pure crystals will be discussed in the next section. The assumptions that we shall make are:

(*a*) There exist on all active slip planes barriers at a mean distance $2L$ from each other; the nature of the barriers will be discussed below.

(*b*) The screw components of the dislocation loops move a distance L' comparable with or greater than L; the screws will be discussed in § 6.

The hardening will be ascribed to the stress round the piled-up groups of n edge dislocations already described. If the surrounding material were elastic, the stress at a distance r from such a group would be

$$nGb/2\pi r \qquad (n \sim 1000).$$

However, at a distance greater than l ($l \sim 10^{-4}$ cm) the material is plastic, as it contains Frank–Read sources. As, however, the stress is extremely unhomogeneous, we do not believe that the dislocations from any source are likely to move a distance much greater than l before meeting a dislocation on another plane and becoming sessile. We therefore have to relieve a local strain $\epsilon(r)$ ($=nb/2\pi r$) by allowing dislocations to move from sources distant l apart through a distance l. It is clear that, at a distance r from the pile-up, each source must contribute $n'(r)$ dislocations, where $n'(r)=l\epsilon(r)/b$, to do this. Round each primary group of n which have moved a large distance L, we have other secondary groups of n' which have moved a small distance l. Moreover, these produce a stress field σ_i at distance r from the primary group having 'wave-length' l and magnitude

$$\sigma_i = Gbn'/l = G\epsilon(r) = Gbn/2\pi r.$$

The magnitude is thus the same as if secondary slip had not occurred.

If, then, a material has slipped so that slip bands are at a distance x apart, the mean distance between the primary groups of piled-up dislocations is $(xL)^{1/2}$.

The mean strain σ_i in the material is then of wave-length l, the dimensions of the network, and of magnitude

$$\sigma_i = nGb/2\pi(xL)^{1/2}. \qquad (9)$$

The strain, on the other hand, is given by

$$\epsilon \sim nb/x. \qquad (10)$$

Eliminating x, we find

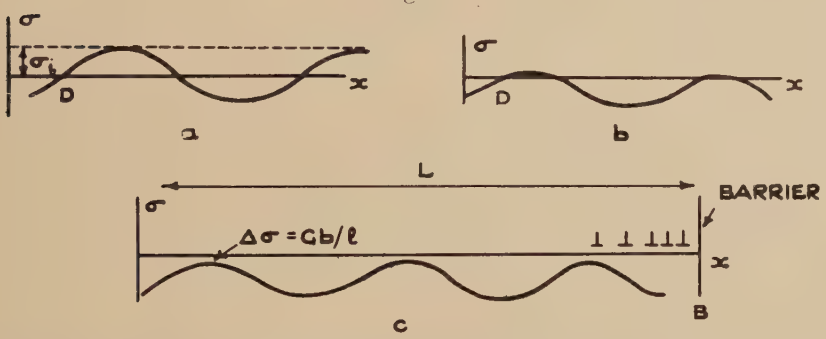
$$\frac{\sigma_i}{G} = \frac{1}{2\pi}\sqrt{\frac{nb}{L}}\,\epsilon^{1/2}. \qquad (11)$$

This formula is to be compared with Taylor's formula (4).

This stress σ_i cannot, as in the original theory of Taylor, be identified with the flow stress at a strain ϵ. The situation is to be pictured as in fig. 5, in which the stress acting on a dislocation is plotted against position. Before an external stress is applied, the situation will be as in fig. 5 (*a*), and the loop of a Frank–Read source will be at or near a point such as D, where the stress vanishes. As the applied stress is increased, the stress within the material will change to that shown in fig. 5 (*b*), the dislocation remaining at D. But a slip band will not be formed until the configuration of fig. 5 (*c*) is reached, the stress at the maximum reaching the value $\Delta\sigma = Gb/l$ required to operate the source. As soon as this is reached, the (dynamic) generation of a slip band occurs ; if there is a barrier at B, the dislocations generated will *all* pile up in its immediate neighbourhood, the mean stress acting on them being $\sigma_i + \Delta\sigma$. But the *number* n of dislocations generated depends on $\Delta\sigma$ only ; for n dislocations near B give at the source a stress field

$$nGb/2\pi L,$$

Fig. 5



Showing internal stress on a material ; dislocations in fig. 5 (*c*) are shown as ⊥.

so that the source will stop generating when

$$nGb/2\pi L = \Delta\sigma ;$$

when this is so,

$$L = nGb/2\pi\Delta\sigma,$$

a formula differing by a factor 2 from (8).

Substituting Gb/l for $\Delta\sigma$, we have

$$nb/L = 2\pi b/l,$$

whence formula (11) for σ_i takes the form

$$\frac{\sigma_i}{G} = \sqrt{\left(\frac{b}{2\pi l}\right)}\,\epsilon^{1/2}. \quad \ldots\ldots\ldots \quad (12)$$

We may take the flow stress to be

$$\sigma_i + Gb/l ; \quad \ldots\ldots\ldots \quad (13)$$

except for small strains the second term is small compared with the first.

Several points of interest may be noted about formula (12). The first is that *the stress-strain curve is independent of the slip distance L.* This is a most important result, because the discussion that we shall give below suggests that the slip distance may depend rather sensitively on the physical conditions. The analysis above shows how the stress-strain curve may not be so dependent.

A second point is that, if we assume that l remains constant throughout the hardening process, formula (12) gives a parabolic stress-strain curve, and if $l \sim 10^{-4}$ cm, it is of the form

$$\sigma/G \sim 100\, \epsilon^{1/2} ;$$

the constant has about the value required for agreement with experiment. In this case, too, it is instructive to write the formula

$$\sigma = \sigma_0 + \sigma_i = \sigma_0 + G\sqrt{\frac{\sigma_0}{G}}\,\epsilon^{1/2}, \qquad \ldots\ldots \quad (14)$$

where σ_0 is the yield stress. It is, however, likely that l will decrease during work-hardening, owing to the processes already described ; in that case the stress will increase as ϵ^m, where m is slightly greater than $\frac{1}{2}$.

One criticism may perhaps be made of this theory. We have to ask why, when the stress is raised, further slip does not occur on planes that have already been active. All the dislocations in the active slip plane, when $\sigma \gg Gb/l$, will have gone to the end ; there are none left in the main part of the slip plane. We suggest that further slip *may* happen on this plane ; but as the random internal stress is constantly changing, there is no reason why the particular centre that has slipped should be more favoured than any other ; and there are always far more sources than are used. In the same way, if the stress is reversed, it will normally be other sources that are active, and not those that have already slipped.

We now turn to a discussion of the nature of the barriers. We remark first that various investigators (Rohm and Kochendörfer 1950, Andrade and Henderson 1951, Lücke and Lange 1952) have observed easy glide in crystals of close-packed metals. The conditions seem to be

(*a*) Very pure materials.
(*b*) Orientation for slip on one plane only.
(*c*) In some cases absence of oxide layer.

By easy glide we mean either the absence of hardening, or slight hardening compared with that normally observed, of magnitude comparable with that in hexagonal metal crystals. Easy glide is not accompanied by the formation of asterisms. It appears that in crystals showing easy glide—as perhaps in hexagonal crystals, the dislocations pass out of the crystals. We believe that when rapid hardening takes place the barriers which stop the dislocations, apart from grain boundaries and oxide layers, are likely to be sessile dislocations, either those already existing or those formed by slip on two planes. If this is so, easy glide on one slip system will occur most easily in very carefully annealed crystals of small total diameter—as in Andrade's work.

This assumption about the nature of the barriers enables us to estimate the slip distance in the absence of other boundaries. If a reasonable proportion of the dislocations in a material are sessile, there should be b/l^3 lines, each of length l, per unit area in any plane. If any one of these is capable of stopping a dislocation in the plane, and if the slip distance L is the same for the screw as for the edge component, we have

$$bL^2/l^3 \sim 1,$$

whence

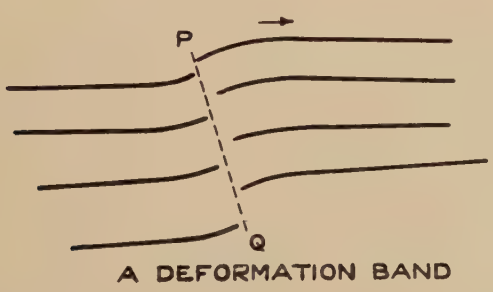
$$L \sim l\sqrt{(l/b)} \sim 10^{-2} \text{ cm}.$$

This gives a value of L of the right order of magnitude, perhaps greater than 10^{-2} cm, since certainly the distance between sessile dislocations will be greater than between the normal ones. Moreover, it gives for n

$$n = \Delta\sigma\, L/bG = L/l = \sqrt{(l/b)},$$

a value which varies with l only as the square root, and may perhaps explain the comparatively constant value of n. On the other hand, conditions which promote flow on two or more sets of planes are bound to

Fig. 6



A DEFORMATION BAND

increase the number of sessile dislocations, and so decrease the slip distance and hence n ; we should certainly expect some decrease in n as the strain is increased. Further experimental work on this point would be of interest.

§5. Deformation Bands

Up to the present we have assumed that all slip bands, except those that reach the surface, terminate in *isolated* groups of piled-up dislocations. In fact, however, there are frequently observed slip lines which terminate in deformation bands ; a schematic representation of a deformation band is shown in fig. 6. Deformation bands are an origin of asterisms, and it has frequently been suggested that they are the origin of hardening.*

* e.g. Mott (1951). Asterisms are not caused by isolated groups of piled-up dislocations. It should be possible in principle to harden a material without forming asterisms.

In our view, however, a group of piled-up dislocations in a deformation band cannot contribute to the hardening at all, because a grid of dislocations or of groups of dislocations of this type has no long-range stress field.* Groups of dislocations which terminate in a deformation band are, from the point of view of work-hardening, wasted.

A deformation band grows in our view in the following manner. The regions P and Q form, owing to the stress field there, very effective traps for dislocations of sign such as to lengthen the band. In a material in which the density of barriers is otherwise low, deformation bands will tend to grow. But unless the density of barriers is very low indeed, there is no reason why a high proportion of slip lines should terminate there.

It is known (Chen and Mathewson 1951, Honeycombe 1951, Laloeuf and Crussard 1951) that deformation bands do not appear when a crystal is stressed under conditions such that slip occurs simultaneously on two sets of planes. On our model, this will be because, with slip on two sets of planes, a large number of sessile dislocations will be formed, leading to a low value of L and inhibiting the growth of deformation bands.

Cottrell (1952) has mentioned another factor which may contribute to strain-hardening when slip on two or more planes occurs. Under such conditions, a dislocation cannot move far without cutting other dislocations and so forming jogs. Energy is required for this, and so an increasingly large stress is required to drive a dislocation through the material. We believe that this cannot be the main cause of hardening ; if it were, then, for impure materials for which there is no question of easy glide, the hardening would depend markedly on the orientation. This is not the case. As Cottrell points out, however, the activation energy required to form a jog is of the order of a only few electron volts even in the absence of a stress. Thus after plastic flow such dislocations as are held up by other dislocations will be easily freed by heat. Cottrell (1953) has given an analysis of logarithmic creep in pure aluminium which suggests that it is due to this process. We believe that Cottrell's mechanism introduces a small temperature-dependent frictional term in the resistance to the motion of dislocations.

§6. Movement of Screw Dislocations and the Generation of Vacancies

The theory of hardening given in the last sections is based on the hypothesis that groups of piled-up edge dislocations can be formed through the agency of the sessile dislocations which results when edge dislocations on intersecting planes coalesce. We have made no mention of the screws, except to assume that their slip distance L' is at least as great as L. We shall in this section discuss the behaviour of the screws ; they are important, because

---

* cf. Nabarro 1952.

(*a*) We believe that moving screw dislocations generate vacancies, and that these are important for the explanation of the dependence of work-hardening on temperature, clustering of slip lines, electric resistance due to cold-work and other effects.

(*b*) In hexagonal crystals or others in which only one slip plane is operative, the blocking of edge dislocations through the formation of sessiles is impossible. We believe that in hexagonal crystals the slip distance L for the edge components must be the same as linear dimensions of the crystal; apart from oxide layers there is nothing to prevent their passing right out of the crystal. The (much slower) hardening which occurs in hexagonal crystals may be associated with piled up groups of screws.

(*c*) A discussion of the behaviour of screws is necessary in order to understand cross-slip.

All these phenomena depend in our opinion on the properties of vacant lattice sites or vacancies, and we shall give a few preliminary facts about these. According to the vacancy theory of diffusion (Johnson 1939), a vacancy is characterized by the energy W required to form it and the activation energy U to move it. In terms of these quantities the coefficient of self-diffusion D is given, apart from a numerical factor, by

$$D=\nu a^2 \exp\{-(W+U)/kT\}.$$

According to the calculations of Huntington and Seitz (1942), W and U should be about equal for a close-packed metal. Thus, if vacancies are formed artificially, e.g. by quenching from a high temperature or by neutron bombardment, self-diffusion is possible at about half the temperature at which it normally occurs. Recent work by Nowick (1951) has confirmed that the exchange of atoms in an alloy (AgZn) is more rapid after quenching, the effect being explained along these lines; Nowick finds for this alloy that, if this hypothesis is correct, U must be somewhat greater than W.

At temperatures at which self-diffusion is possible, the absorption or generation of vacancies enables an edge dislocation to 'climb', i.e. to move in a direction perpendicular to its slip direction. As described by the present author (Mott 1951), such a process is almost certainly responsible for the polygonization of a deformation band.

There is much evidence that *moving* dislocations generate vacancies at temperatures too low for them to be formed thermally. The evidence has been fully reviewed by Seitz (1952). The two main facts quoted by Seitz are:

(*a*) Cold-work increases the conductivity of alkali-halides, a conductivity thought to be ionic and to be due to vacancies. After cold-work the resistance gradually increases at a rate consistant with this hypothesis.

(*b*) The increased electrical resistance of metals due to cold work can be partly or mainly removed by heating to a temperature much below that at which the mechanical 'recovery', or drop in the flow stress, occurs. It is thought that a large part of the electrical resistance is due to vacancies.

formed by the cold-work; this will disappear if the metal is raised to a temperature at which they can move; while to get rid of mechanical hardness one has to allow edge dislocations to ' climb ' out of the piled-up groups (Mott 1951), which will occur only if the specimen is heated to near to the temperature of self-diffusion.

Recent work at Harwell (Dugdale 1952) has given striking confirmation of the correctness of this hypothesis about the origin of the residual resistance due to cold-work. Platinum wires have been exposed to fast neutron bombardment in the pile, and an increase in the electrical resistance of the order of $\frac{1}{2}\%$ observed. It is found that this resistance begins to diminish on annealing at about 100° c, and the activation energy for this process was determined, and found to be about 1·1 ev. Specimens with a residual resistance due to work-hardening were annealed in the same way and for these specimens too the activation energy for resistance decrease was found to be 1·1 ev. Since in the first case the resistance was almost certainly due to vacancies and interstitial atoms and the annealing involved the movement of the former, it seems highly probable that this is true also for the cold-worked specimens.

We turn now to the mechanism by which vacancies are formed by moving dislocations. Seitz (1952) gives a review of the various possibilities; of these we consider the following the most likely.

Consider the movement of dislocations in a given slip plane. Suppose that a certain number of screw dislocations cut this plane; in this context by a screw dislocation we mean one with Burgers vector *not* in the slip plane considered.* We shall call such dislocations ' crossing dislocations ' and the points where they cut the slip-plane under consideration ' crossing points '. Now, whenever a moving dislocation passes a ' crossing point ', a jog is generated in that dislocation; and although a jog once formed is no impediment to the motion of an edge dislocation, a screw dislocation with a jog in it cannot move without leaving behind it a row of vacant lattice sites *or* interstitial atoms, according to the sign. This geometrical result of the dislocation concept was first put forward by Thornton Read, and is fully discussed by Seitz (1952) and by Mott (1951).

We may thus accept the hypothesis that moving screw dislocations do produce rows of vacancies and interstitial atoms, and that this process means that there is a resistance to the motion of screws which is not shared by edges. However, it is not immediately obvious that the passage of a *group* of dislocations along the slip plane, as in the formation of a slip band, will produce the vacancies and interstitials that we require. For, as pointed out by Mott (1951), provided the crossing screws do not move during the formation of a slip band, the rows of atoms added or subtracted by each succeeding dislocation will be side by side; all that will result from the motion of n screws a distance L' beyond a crossing

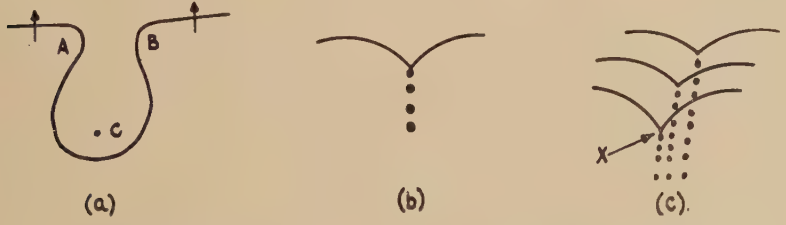
* In the simple cubic structure such a dislocation *must* have a screw component, and this is also the case for hexagonal structures, but not for the close-packed cubic.

point will be the removal or addition of a slab of atoms of width nb and length L'. After the passage of the first few dislocations, the passage of further ones will meet with little resistance.

Seitz has suggested that, by a mechanism to be described below, enough energy is *released* in the formation of vacancies to disperse them. If this is true for the whole slip distance L' (say 10^{-2} cm), far too many vacancies are formed and the resistance to the motion is much too large to be compatible with the observed facts. In order to explain the observed resistance, Seitz has to assume that a moving dislocation forms one vacancy for every 10^4 atoms that it crosses. If $1/l^2$, the density of crossing points, is such that $l \sim 10^{-4}$ cm, this will be the case after a dislocation has moved a distance of the order l ; after it has moved a distance L' it will have jogs at intervals of l^2/L' along its length, and, if L' is 10^{-2} cm, will be creating vacancies at 100 times the observed rate.

Moreover, if x is the distance between jogs, the resistance to the motion of a screw is equivalent to a stress of order Gb/x, and only if x is greater than l is this comparable with or less than the yield stress.

Fig. 7



Motion of a screw dislocation cutting another screw.

We believe that the correct model is a compromise between the two points of view. Each screw dislocation, when it approaches a crossing point C, is repelled by the crossing dislocation, and curls round it, as in fig. 7 (*a*). When the two edge components, now on contiguous planes, join together and form a row of vacancies, enough energy is released to disperse them. Thereafter, however, and thus at a distance of order l from C, a continuous screw line is formed as in fig. 7 (*b*), which leaves behind it a row of vacancies which are not dispersed. The *first* process generated about l/b dispersed vacancies every time a screw passes a crossing point ; the *second* generates rows of vacancies next to each other, and thus removes *layers* of atoms, presenting no effective resistance to the movement of dislocations after the first few.

Such a mechanism accounts for the generation of about the right number of vacancies ; we have to ask further

(*a*) Can we give any explanation of the limited slip along one plane, similar to that provided by sessile dislocations in the case of edges ?

(*b*) Can we find a mechanism which prevents the screws from moving back under a reversed stress ?

(*c*) Can we account for the (slow) work-hardening of hexagonal single crystals?

As regards (*a*) we suggest the following: if n screw dislocations have moved a distance L, for each of the L/l^2 crossing dislocations cut per unit length, a plane of atoms of width nb is either added or removed. The crossing dislocations are pushed sideways, therefore, a distance nb, for which energy is required to increase their length. We may suppose that the resultant force pushing the n screws back is equivalent to a stress

$$GnLb^2/l^3.$$

Which increases with increasing n.

The model suggests, then, that the slip on each band increases with increasing strain. That the slip on each plane increases with increasing strain is in accordance with the observations of Andrade and Roscoe (1937) on zinc.

The model will also lead us to expect that as the strain is increased the steps on the surface of a cubic metal may lengthen, though they do not increase in height (at low temperatures).

With regard to the sticking or anchoring of screw dislocations, we suppose that they will always have a certain number of 'free jogs'~i.e. jogs which are not anchored to a point such as X in fig. 7 (*c*). These, and also any movement of the crossing screws since the slip bands were formed, will hinder movement in either direction.

Finally, we discuss the work-hardening of hexagonal crystals (or of cubic crystals under conditions of easy glide). This may by essentially of the same nature as in cubic crystals, the screws replacing the edges. We suppose that in hexagonal crystals the edges escape to the surface, so that only the screws are stuck, the worked crystal being full of screws. The slow rate of hardening of hexagonal crystals will be due primarily to the much greater tendency that one would expect, there being no possibility of sessile screws, for the groups of dislocations to form into grids, with the result that there is no *long-range* stress from any group. Some evidence that this occurs is presented in the next section (on cross-slip). We have not, however, been able to formulate any quantitative theory for these crystals.

An alternative suggestion is that in hexagonal crystals and in cubic crystals under conditions of easy glide the screws escape as well as the edges, and that the hardening is simply due to the strains round vacancies formed.

§ 7. Cross-slip

Up to this point we have implicitly assumed that dislocation loops remain in the plane in which they start to move. If this were the case, slip bands observed on the surface of metals would be straight. Actually, in cubic metals, the slip bands observed frequently show intimate cross-slip. Examples of cross slip have been given by Cahn (1951). The points to note are:

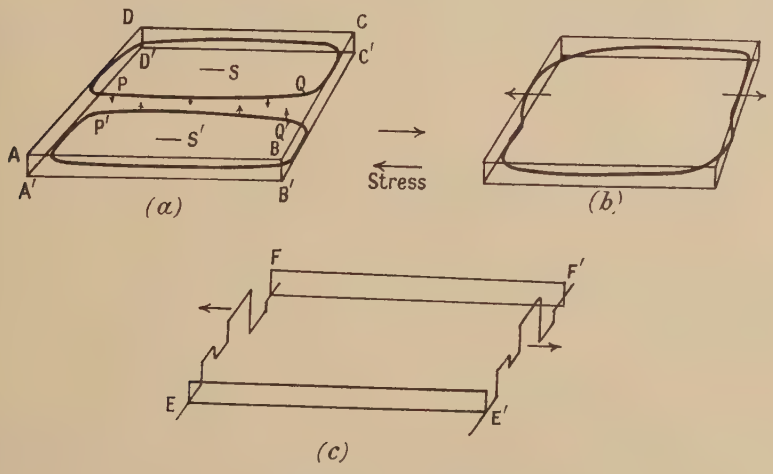
(i) Slip bands on the (111) planes on which the stress is largest are connected by short bands on the other sets of (111) planes.

(ii) Cross slip is more marked at high temperature.

(iii) The photographs at high magnification show that the step connecting the main bands is less sharp than the main steps.

An explanation of cross-slip (due to Frank) was given in the author's Guthrie lecture (Mott 1951). This explanation depends on the different kinematical properties of the edge and of the screw dislocation. The edge dislocation, if it is straight, can move only in the plane defined by its Burgers vector and itself. Moreover each element of an edge dislocation line which switches from one (111) plane to another has this property too. The same is true of dislocation lines which are partially edge, partially screw. A pure screw dislocation, on the other hand, has its Burgers vector parallel to itself, and so is not associated with any particular one

Fig. 8



Model for cross-slip.

of the (111) planes in which the Burgers vector lies. There is no kinematic reason, therefore, why a screw dislocation should not move for a certain distance in one of the (111) planes and then switch to another.

The explanation of cross-slip already referred to envisages, then, two loops spreading out from sources, marked S, S′ in fig. 8, in parallel planes not too far from each other. PQ, P′Q′ are the screw elements. When PQ reaches a position above P′Q′ the two lines having opposite sign attract each other. They can then leave the planes in which they were moving originally and move towards each other until they join and annihilate each other. One is then left with a ring in which the two edge elements switch from one plane to another (fig. 8 (*b*)). This may happen for a number of sources, so that finally one is left with the two kinky edge

dislocations of fig. 8 (*c*), connected by screws. When these edge dislocations appear on the top surface of the crystal, and are followed by others produced in the same way, they will form kinky steps of the type observed.

This model shows why the steps joining the main slip bands are less distinct ; when one loop after another starts out from the sources S, S′, there is no obvious reason why the screws should move from PQ to P′Q′ always along the same plane.

We believe that this description is geometrically correct. It needs, however, some amplification. One may ask, for instance :

(*a*) Why should the sources SS′ be generating loops at the same time ? They will not in general have the same values of l and σ_i and one would expect one source to generate its full complement of loops before the next was brought into play (cf. Crussard 1951).

(*b*) In close-packed cubics—as opposed to the simple cubic, a very considerable activation energy will be required to make a screw dislocation moving in one plane switch to another. This is because, as already stated, all dislocations, screw or edge, split into half dislocations.

Fig. 9

Showing activation process for the motion of an extended screw dislocation out of its plane, the dotted lines being in new plane. Each line represents a half dislocation.

The half screw dislocation does not have its Burgers vector parallel to itself ; it cannot move out of its plane. If a screw is to move out of its plane, a small length must be moved out first, and at the points where the two join, the two halves must be brought together, as in fig. 9. The necessary activation energy will be several electron volts, probably too large for the process to occur.

We believe that the clue to both criticisms is to be found in another property of the model which has already been emphasized. A screw dislocation, after it has cut a certain number of other screws, contains jogs, so that the resistance to the motion of a screw will increase as it moves along, and other screws will pile up behind it. The stress induced by this pile-up will readily induce some nearby source to generate loops, so that one is left with the configuration shown in fig. 10. Slip has occurred from A to B and A′ to B′ ; energy will be gained if the piled-up screws at B, B′, can move into the plane BB′ (shown dotted) and so annihilate each other. The experimental evidence shows that this only occurs when the temperature is raised ; we suggest that the clue is the

high stresses due to the pile-up ; these either allow spontaneous generation of loops in the plane BB′, or much reduce the activation energy for the process of fig. 9.

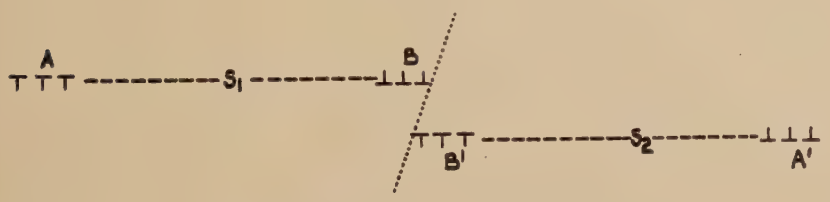
In a recent paper Smith (1952) has shown that pure copper bars extended by *c.* 10% show a drop in Young's modulus of about 10%, but that this is recovered on annealing at 100° c. This may well be due to the motion of screw dislocations bounding the slip bands (the edges being locked), and the process by which annealing takes place exactly the same as cross slip, namely the mutual annihilation of the screws.

§8. Dependence of Work-hardening on Temperature

The theory of work-hardening of cubic crystals given in §§ 2, 3 and 4 is applicable as it stands to work-hardening at low temperatures only. In this section we shall suggest explanations of the following phenomena which depend on temperature:

(*a*) The ' clustering ' of active slip planes at distances of the order of 200 Å from each other observed in aluminium at room temperature and above and in other metals, but absent according to Brown (1951) in aluminium at the temperature of liquid air.

Fig. 10



Showing groupings of screw dislocations.

(*b*) The dependence of the whole stress-strain curve on temperature.

(*c*) The behaviour observed when a crystal is strained at room temperature and then strained further at a lower temperature.

We remark first that the theoretical values of all the flow stresses derived in this article are linear functions of G, the shear modulus, and that G itself depends on temperature ; for aluminium the change is of the order 7% per 100° c (Ke 1947) above the characteristic temperature. Thus changes of from 10 to 20% between the absolute zero and room temperature may be expected from this cause alone. Apart from this, one expects very little temperature-dependence for a yield stress determined by a relation of the type

$$\sigma_0 = Gb/l.$$

The activation energy necessary to operate a Frank–Read source for stress σ less than σ_0 is (Mott and Nabarro 1948)

$$k\sigma_0 bl^2(1-\sigma/\sigma_0)^{3/2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where k is a numerical factor of order 0·2. The term $\sigma_0 bl^2$ is equal to $Gb^3(l/b)$ and thus to *c.* 10^4 ev. If, then, a yield point with a dependence on temperature greater than G is observed, as for instance in Rosi and Mathewson's (1950) work on aluminium, the most probable conclusion is that the yield point is determined by Cottrell locking.*

Another process in which a large dependence on temperature may be expected is the spontaneous formation of dislocation loops in regions of very high stress ; this has already been mentioned in connection with cross-slip. Another is Cottrell's mechanism of the cutting of one dislocation by another ; this we believe responsible for transient creep at low temperatures.

Apart from these and the variation of G with temperature, it seems impossible to account for any temperature dependence of work-hardening in terms of an activated motion of dislocations ; the term l in formula (15) should be at least 10^{-4} cm or possibly even comparable with the slip distance. We consider, therefore, that the dependence on temperature is largely to be accounted for through the action of vacancies.

We assume that thermal recovery is associated with the ' climb ' of dislocations out of their slip planes, a process possible at temperatures for which self-diffusion can occur ; a theoretical account has been given by the author (Mott 1951). Such a process enables the piled-up arrays of dislocations, to which we have ascribed work-hardening, to disperse and deformation bands to polygonize. We suggest, however, that at considerably lower temperatures, namely those at which vacancies can move, the vacancies created by the moving screw dislocations will soften the material in the immediate neighbourhood of the planes which have already slipped.† This could occur through the diffusion of vacancies to the *secondary* groups of piled-up dislocations, the existence of which has already been deduced. The temperature above which this can occur will be *c.* $-70°$ c in aluminium, room temperature in copper, *c.* 120° c in nickel.‡

If this occurs we have, following Brown, an immediate explanation of clustering. Once slip has occurred on a given plane there will occur a relatively slow reduction of σ_i in the immediate neighbourhood of that plane. Slip will then occur again, either on the plane which has slipped already or on some neighbouring one ; but the source that has generated vacancies already is in no way preferred ; in most cases it will be another one. The nearest of these should be at a distance from the first plane l^3/LL' ; with $L \sim L' \sim 10^{-2}$ cm and $1/l^3 \sim 10^{10}$ cm^{-3}, this gives 10^{-6} cm.

---

* Note that Neurath and Koehler (1951) find almost the same yield point of pure single crystals of copper at liquid air and room temperature.

† Brown (1951) has suggested, without proposing a mechanism, that the material round a slip plane anneals, and has explained clustering and certain other phenomena in terms of this assumption. The discussion given here owes much to Brown's work.

‡ The basis for this estimate is given below.

We thus expect sources at about this distance to become active. The lines of a cluster, on this hypothesis, will be similar in nature to those of the primary slip, and will, like the primary slip band, have groups of piled-up edge dislocations at the extremities. We suppose, moreover, that the limited supply of vacancies affects the secondary rather than the primary groups of piled-up dislocations.

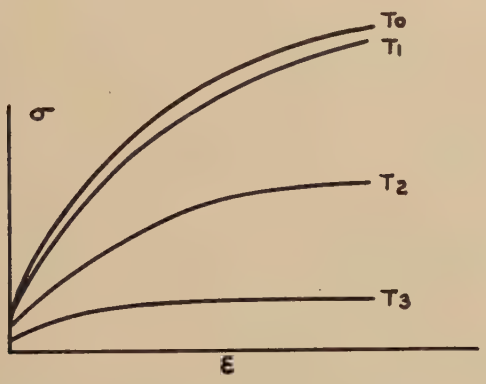
If this hypothesis is correct, we may, following Brown, account for the temperature-dependence of the work-hardening curve. At temperatures T_0, T_1 (fig. 11), too low for vacancies to move, we shall expect a *small* change in σ proportional to that for G. At a temperature T_2, at which vacancies can move, we shall expect :

(*a*) A rather rapid drop of the whole curve, it being assumed that most new slip bands start near old ones.

(*b*) The disappearance of the main part of the residual resistivity produced by cold-work.

(*c*) The formation of clustered slip lines.

Fig. 11



Work-hardening at temperatures such that $T_3 > T_2 > T_1 > T_0$.

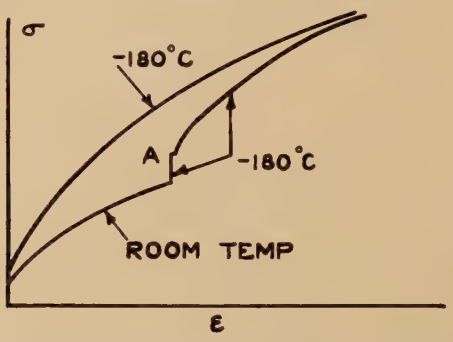
At a temperature T_3 more than twice as high, true recovery may occur during straining.

Finally, following a suggestion of Brown, we can explain with this model the behaviour of the stress-strain curve of a crystal such as aluminium strained at room temperature and then lowered to liquid air temperature (fig. 12). On cooling to liquid air temperature there is a small jump in σ, doubtless due to the change in G ; then hardening is very rapid until the curve characteristic of liquid air temperature is approached. Now in our model the quantity n/L—i.e. the *strength* of the groups of piled-up dislocations divided by the distance between them, will not depend on the temperature of deformation ; moreover it is this quantity, according to formula (11), which determines the flow stress. The crystal at room temperature flows more easily because the material round any slip plane softens, owing to a rearrangement of the secondary groups. Thus when slip

continues from the point A in fig. 12, the annealed region is quickly used up ; thereafter the material differs from a crystal strained from the beginning at liquid air temperature only in that some of the slip bands are grouped in clusters, but not in the total number of dislocation groups or in the mean value of n/L.

A further consequence of the formation of vacancies is probably, as has been suggested by various authors (Hall and Williamson 1951), the fragmentation or substructure that is formed in a cold-worked material (Gough and Wood 1936, 1938). The elements of the substructure are, we suggest, simply polygonized crystallites, and polygonization is enabled to take place through the movement of vacancies formed by the cold work. This suggests that fragmentation would not take place at a low enough temperature.

Fig. 12



Stress strain curve initially at room temperature and then at liquid air temperature.

Finally, we give a very rough estimate of the critical temperatures T_1 below which the vacancies cannot move. Unfortunately we have only one definite measurement of the activation energy U for movement of a vacancy, that of Dugdale (1952) for platinum, and this is less than one half the probable value of W for this element, instead of being equal to it. The following table gives $U+W$, the activation energy for self-diffusion for the elements for which it is known, and for Al, Pt, Ni deduces it on the assumption that these quantities are proportional to the melting points. U is deduced on the assumption that the ratio U to W is the same as for platinum. Finally T_1 is taken, rather arbitrarily, as the temperature at which a vacancy makes one jump in ten seconds.

Commenting on these results, we may note that Seitz (1952, p. 45) reports unpublished experiments by Blewitt which record that the resistance of copper strained at liquid air temperature shows a partial recovery on annealing at $-80°$ c ; with the above value of U, we may estimate that at this temperature a vacancy would change places about once in 1 000 sec, so an anneal at this temperature is just possible. It is difficult on this basis

to understand the results of Bowen, Eggleston and Kropschott (1952) who after cold work at room temperature found an activation energy of 28·3 kcal for further recovery of the residual resistance of copper. Perhaps this is the activation energy for the motion of interstitial atoms.

Table 1

Metal	$W+U$	U	T_1 (degrees K)
Cu	**46·5**	16	230
Ag	**45·9**	16	230
Au	**53·0**	18	260
Pb	**27·5**	9·5	140
γ-Fe	**67·9**	23	350
Al	39	13	200
Pt	75	**26**	430
Ni	64	23	390

W and U are in kcal/g atom.
Experimentally determined quantities in bold-face type.

References

Andrade, E. N. da C., and Henderson, C., 1951, *Trans. Roy. Soc.*, **244,** 177.
Andrade, E. N. da C., and Roscoe, R., 1937, *Proc. Phys. Soc.*, **49,** 152.
Bowen, D., Eggleston, R. R., and Kropschot, R. H., 1952, *J. Applied Phys.*, **23,** 630.
Brown, A. F., 1949, *Nature, Lond.*, **163,** 961 ; 1951, *J. Inst. Metals*, **80,** 115 ; 1952, *Advances in Physics*, **1** (in press).
Brown, A. F., and Honeycombe, R. W. K., 1951, *Phil. Mag.*, **42,** 1146.
Cahn, R. W., 1951, *J. Inst. Metals*, **79,** 129.
Cottrell, A. H., 1948, *Bristol Conference on the Strength of Solids, Phys. Soc.*, p. 30 ; 1952, *Phil. Mag.*, **43** (in press) ; 1953, *Dislocations and Plastic Flow in Solids* (Oxford : Clarendon Press) (in press).
Cottrell, A. H., and Bilby, B. A., 1949, *Proc. Phys. Soc.* A, **62,** 49.
Chen, N. K., and Mathewson, C. H., 1951, *J. of Metals*, **3,** 653.
Chen, N. K., and Pond, R. B., 1952, *Phys. Rev.* (in press).
Crussard, C., *Solvay Conference* 1951, discussion remarks p. 575.
Dugdale, R. A., 1952, *Phil. Mag.*, **43,** 912.
Eshelby, J. D., 1949, *Proc. Roy. Soc.* A, **197,** 396.
Eshelby, J. D., Frank, F. C., and Nabarro, F. R. N., 1951, *Phil. Mag.*, **42,** 351.
Fisher, J. C., Hart, E. W., and Pry, R. H., 1952, *Phys. Rev.* (in press).
Frank, F. C., 1951, *Pittsburgh Conference on Plastic Deformation of Crystalline Solids*, p. 89 ; 1952, *Advances in Physics*, **1,** 91.
Frank, F. C., and Read, W. T., 1950, *Phys. Rev.*, **79,** 722.
Gough, H. J., and Wood, W. A., 1936, *Proc. Roy. Soc.* A, **154,** 510 ; 1938, *Ibid.*, **165,** 358.
Hall, W. H., and Williamson, G. K., 1951, *Proc. Phys. Soc.* B, **64,** 937.
Heidenreich, R. J., and Shockley, W., 1948, *Bristol Conference on the Strength of Solids*, p. 57.
Herring, C., and Galt, J., 1952, *Phys. Rev.*, **85,** 1060.
Honeycombe, R. W. K., 1951, *J. Inst. of Metals*, **80,** 45.
Huntington, H. B., and Seitz, F., 1942, *Phys. Rev.*, **61,** 315.
Johnson, R. P., 1939, *Phys. Rev.*, **56,** 814.
Ke, T. S., 1947, *Phys. Rev.*, **71,** 533.

Laloeuf, A., and Crussard, C., 1951, *Revue de Metallurgie*, **48,** 461.
Leibfried, G., 1949, *Z. Physik*, **126,** 781 ; 1950, *Ibid.*, **127,** 344.
Lomer, W. M., 1951, *Phil. Mag.*, **42,** 1327.
Lücke, K., and Lange, H., 1952, *Z. für Metallkunde*, **43,** 55.
Masing, G., and Raffelspeiser, J., 1950, *Z. für Metallkunde*, **41,** 65.
Mott, N. F., 1951, *Proc. Phys. Soc.* B, **64,** 729.
Mott, N. F., and Nabarro, F. R. N., 1948, *Bristol Conference on the Strength of Solids*, p. 11.
Nabarro, F. R. N., 1951, *Proc. Roy. Soc.* A, **209,** 278 ; 1952, *Advances in Physics*, **1,** 269.
Neurath, P. W., and Koehler, J. S., 1951, *J. Applied Phys.*, **22,** 621.
Nowick, A. S., 1951, *Phys. Rev.*, **82,** 551.
Röhm, F., and Kochendörfer, A., 1950, *Z. für Metallkunde*, **41,** 265.
Rosi, F. D., and Mathewson, C. H., 1950, *J. of Metals*, **188,** 1159.
Seitz, F., 1952, *Advances in Physics*, **1,** 43.
Smith, A. D. N., 1951, *M.O.S. Report*, to be published.
Stroh, A. N., 1952, *Proc. Phys. Soc.* (in press).
Taylor, G. I., 1934, *Proc. Roy. Soc.* A, **145,** 362.

CXVIII. *The Propagation of a Radio Atmospheric—II.*

By K. G. Budden
Cavendish Laboratory, Cambridge*

[Received August 19, 1952]

Contents

Summary

In a previous paper (Budden 1951 a, referred to as Paper I), it was shown that the propagation of an atmospheric could be treated either in terms of the successive reflection of rays at the earth and the ionosphere, or by considering the space between the earth and the ionosphere as a wave-guide, and discussing the properties of the various wave-guide modes. An outline of the theory was given for the case when the top surface of the wave-guide was the boundary of a homogeneous ionized medium, and the earth's magnetic field was neglected.

The present paper gives the full mathematical theory, and shows how the characteristics of the wave-guide modes may be determined in the most general case, provided that the reflecting properties of the earth and the ionosphere are known as functions of the angle of incidence which is, in general, complex. It is found that the wave-guide modes

* Communicated by J. A. Ratcliffe, F.R.S.

are of two types, which are described as ' quasi-transverse magnetic ' and ' quasi-transverse electric '. Allowance can be made for the curvature of the earth.

A particular case is then discussed, in which the surface of the earth is assumed to be perfectly conducting, and the ionosphere is assumed to be a homogeneous ionized medium, with the steady magnetic field of the earth superimposed. The results of numerical calculations are given, for a few special cases, in the form of curves. These show (i) the attenuation in the various modes as a function of frequency, (ii) the polarization characteristics of the wave in typical modes, and (iii) the amplitudes of the waves which are excited in typical modes by a vertical electric dipole source.

§1. Introduction

In a previous paper (Budden 1951 a, hereinafter referred to as Paper I), it was shown that the propagation of an atmospheric could be treated either in terms of the successive reflection of rays at the earth and the ionosphere, or by considering the space between the earth and the ionosphere as a wave-guide, and discussing the properties of the various wave-guide modes. The method of wave-guide modes was shown to be the simpler for propagation to great distances, since there are then only one or two modes of large amplitude, and the modes of higher order are heavily attenuated. The characteristics of the various modes were evaluated by a method which was essentially equivalent to Eckersley's phase integral method (Eckersley 1932), and a physical argument was advanced to justify the use of this method. A mathematical justification of the method was given only in outline in Appendix A of Paper I. The effects of the earth's magnetic field and of the earth's curvature were neglected.

The object of the present paper is to give a fuller mathematical treatment of the same problem, and to show how the theory can be extended to include the effects of both the earth's curvature and the earth's magnetic field. As in Paper I, the ionosphere is still treated as homogeneous and sharply bounded, but is no longer regarded as isotropic.

§ 2. Some General Theoretical Ideas

2.1. *The Reflection of Very Long Radio Waves at the Ionosphere*

Recent experiments on the propagation of very long radio waves from commercial radio transmitters have shown that reflection of these waves at the ionosphere is accompanied by a change of polarization. (See for example, Bracewell *et al.* 1951.) It is convenient to describe the incident and reflected waves in terms of component electric fields parallel to, and perpendicular to, the plane of propagation,* and to consider four co-

* The component in the plane of propagation is sometimes called the ' normal component ', and that perpendicular to the plane of propagation the ' abnormal component '.

efficients denoted by ${}_{\parallel}R_{\parallel}$, ${}_{\parallel}R_{\perp}$, ${}_{\perp}R_{\parallel}$, ${}_{\perp}R_{\perp}$, to indicate the complex ratio of a specified electric field in the wave after reflection to a specified electric field in the wave before reflection. The first subscript denotes whether the electric field specified in the incident wave is parallel ($\parallel$) or perpendicular ($\perp$) to the plane of incidence, and the second subscript refers in the same way to the electric field in the reflected wave.

The set of four conversion coefficients can be written in matrix form thus:

$$\mathbf{R}=\begin{bmatrix} {}_{\parallel}R_{\parallel}, & {}_{\perp}R_{\parallel} \\ {}_{\parallel}R_{\perp}, & {}_{\perp}R_{\perp} \end{bmatrix}. \qquad \ldots\ldots\ldots \quad (1)$$

In this paper, as in Paper I, the ground will be treated as a perfect conductor. For reflection at the ground, the reflection coefficient matrix therefore becomes

$$\mathbf{R}_g=\begin{bmatrix} 1, & 0 \\ 0, & -1 \end{bmatrix}. \qquad \ldots\ldots\ldots \quad (2)$$

If we consider two successive reflections, the first at the ground, and the second at the ionosphere, the pair of reflections is together represented by the matrix

$$\mathbf{R}_2=\begin{bmatrix} {}_{\parallel}R_{\parallel}, & -{}_{\perp}R_{\parallel} \\ {}_{\parallel}R_{\perp}, & -{}_{\perp}R_{\perp} \end{bmatrix}. \qquad \ldots\ldots\ldots \quad (3)$$

It is readily shown that any number of successive reflections is represented by a matrix which is the product of the matrices appropriate to individual reflections, multiplied in the correct order by the ordinary rule for multiplication of matrices.

2.2. *Sources and Image Sources*

In Paper I the source of an atmospheric was assumed to be a vertical electric dipole, whose dipole moment varied with time in a known way. The dipole moment was Fourier analysed into sinusoidally varying components, and the propagation of the component frequencies was considered separately. In finding the impulse response of the waveguide system, a Fourier integral was evaluated which included a factor representing the propagation characteristic of the system as a function of frequency. In the present paper we confine our attention to one component frequency of the source. The source is therefore considered to be an oscillating vertical electric dipole of angular frequency p. If the impulse response of a system is required, it can be found by evaluating the Fourier integral which contains the transmission characteristic as a function of p.

The signal at a point distant from the source can be considered as made up of rays travelling direct from the source, together with rays reflected one or more times at the ionosphere and at the ground. It is well known that these rays could be produced by an array of point sources placed at the positions of the images of the source as seen by

reflection in the two surfaces. If the source is at the ground, and the ionosphere is at height h, this array of images acts like a diffraction grating to give spectra at angle i with the vertical given by:

$$2h \cos i = N\lambda, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where N is an integer, and λ is the wave-length.

The zero order spectrum forms the zero order mode in the wave-guide. The two spectra for $N=+1$, and $N=-1$, may be regarded as the component waves of the mode of order 1. If now the ionosphere is taken to be an imperfect reflector, the image dipoles are no longer equal, but become weaker as we go further away from the central one. The images correspond to two diffraction gratings, one above and one below the ground, in each of which the sources are not all of equal intensity. A series of spectra is formed as before, but in each spectrum the energy is now spread over a small range of angles.*

The following way of considering the spectrum of one order, produced by the grating of unequal sources, is also useful. It is possible to find a complex angle, i_N, for each spectrum such that the energy in the spectrum may be considered to be confined entirely to a wave whose normal is inclined at this complex angle to the vertical. Consider the contributions to the Nth order spectrum of two adjacent points in one of the gratings. These contributions consist of two plane waves whose (complex) phase difference is $2h \cos i_N$. The ratio of the two contributions is given by:

$$R_2 \exp(-2jkh \cos i_N). \quad . \quad . \quad . \quad . \quad . \quad (5)$$

R_2 is a complex number representing the effect of two reflections, one at the ground, and one at the ionosphere. If the ionosphere is isotropic, R_2 is a pure number, but we shall later discuss the case where R_2 is a matrix. By analogy with an ordinary grating, having sources of equal amplitude, we may infer that the components from adjacent points of a grating must agree in both amplitude and phase, and the expression (5) must therefore be unity. We then have:

$$\exp(2jkh \cos i_N) = R_2 \exp(2\pi jN). \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This is the same as eqn. (3) of Paper I.

Now consider the field which results from combining the spectrum of order $+N$ from the top grating and that of order $-N$ from the bottom grating. From symmetry it is clear that the resolved parts of the propagation constants in the horizontal direction for these two spectra are the same. In fact these spectra are the upgoing and downgoing component waves of the wave guide mode of order $+n$. It is clear on this picture why the properties of modes of order $+n$ and $-n$ are different.

* The situation is similar to that which occurs with an optical Lummer–Ghercke plate. The finite angular spread in each spectrum corresponds to the finite resolving power of the plate.

Modes of one sign relate to spectra of the gratings for which the wave normal is inclined towards the direction of increasing amplitude of the grating points, and those of the other sign to spectra with the wave normal inclined towards the direction of decreasing amplitude. It may at first be difficult to see how an array of images above the top boundary of the guide can give rise to an upgoing wave (real part of cos i_N positive) below the boundary. Such an upgoing wave must, however, exist, to satisfy the boundary conditions. A fuller mathematical treatment of the problem is given in §§ 3.1 to 3.3.

2.3. *The Hertz Vectors*

The vertical dipole source will be assumed to have a dipole moment $P_{\parallel} \exp(jpt)$. The field of such a dipole is conveniently represented by the electric Hertz vector $\mathbf{\Pi}_{\parallel}$. This is everywhere parallel to the dipole source, and its magnitude is a scalar whose value at any point in a medium of dielectric constant ϵ and unit permeability is

$$\Pi_{\parallel}=P_{\parallel} \exp(jpt) \exp(-jkr)/r, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where r is the distance from the point to the source, and $2\pi/k$ is the wavelength in the medium. The electric and magnetic fields of the dipole are then given by

$$\mathbf{E}_{\parallel}=(1/\epsilon_0\epsilon) \text{ grad div } \mathbf{\Pi}_{\parallel}-\mu_0\partial^2\mathbf{\Pi}_{\parallel}/\partial t^2, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\mathbf{H}_{\parallel}=\partial/\partial t \text{ (curl } \mathbf{\Pi}_{\parallel}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where ϵ_0 and μ_0 are the electric and magnetic permittivities of free space. If the waves reflected at the earth and the ionosphere have no abnormal component, then the image dipoles are all vertical electric dipoles, and the fields due to them can also be described by electric Hertz vectors. The full expression for $\Pi_{\parallel}$ is then a series of terms representing the contributions from the various image dipoles. Such a series was given in Appendix A of Paper I.

If the first reflected wave from the ionosphere includes some abnormal component, we may consider this as coming from the same image point as that which gives rise to the normal component, but the image source must now include a part which gives rise to a horizontal electric field. This part must have the same polar diagram as the source dipole. An image with these properties is a vertical magnetic dipole. Each image therefore consists of an electric and a magnetic dipole superimposed. The field of a vertical magnetic dipole $P_{\perp} \exp(jpt)$ at the origin is conveniently described by the magnetic Hertz vector $\mathbf{\Pi}_{\perp}$ which is always parallel to its source. Its magnitude is given by

$$\Pi_{\perp}=c\epsilon_0 P_{\perp} \exp(jpt) \exp(-jkr)/r.$$

The factor $c\epsilon_0$ is included so that $\Pi_{\parallel}$ and $\Pi_{\perp}$ have the same dimensions. The associated electric and magnetic fields are given by

$$\mathbf{E}_{\perp}=-\mu_0 c\partial/\partial t(\text{ curl } \mathbf{\Pi}_{\perp}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\mathbf{H}_{\perp}=c \text{ grad div } \mathbf{\Pi}_{\perp}-(\epsilon/c)\partial^2\mathbf{\Pi}_{\perp}/\partial t^2, \quad . \quad . \quad . \quad . \quad (11)$$

where c is the velocity of light in free space. The complete field at any point is therefore described by the single column matrix

$$\mathbf{\Pi}=\begin{pmatrix}\Pi_{\parallel}\\ \Pi_{\perp}\end{pmatrix}. \qquad \ldots\ldots\ldots \quad (12)$$

The field in each component wave of any mode can also be represented by a matrix of this form, and the field after reflection of such a component wave is given by the matrix product: $\mathbf{R\Pi}$.

The contribution to $\mathbf{\Pi}$ from a vertical electric dipole is

$$\mathbf{P}\{\exp(jpt-jkr)\}/r, \qquad \ldots\ldots\ldots \quad (13)$$

where

$$\mathbf{P}=\begin{bmatrix}P_{\parallel}\\ 0\end{bmatrix}. \qquad \ldots\ldots\ldots \quad (14)$$

2.4. *The Wave-Guide Modes*

The space between the earth and the ionosphere may be regarded as a wave-guide of infinite horizontal width, but of finite vertical depth h. If the ionosphere is isotropic, then in such a guide modes of two types may be propagated, usually described as TM and TE modes. In the TM modes the magnetic field is everywhere horizontal, and the electric vector is entirely in the plane of propagation. In the TE modes, the electric field is horizontal, and the magnetic field is entirely in the plane of propagation. Only TM modes can be excited by a vertical electric dipole and hence in Paper I only TM modes were discussed. Any TM mode has an order n, which may be a positive or negative integer or zero. The zero order TM mode played an important part in the theory of propagation of atmospherics. If the earth and the ionosphere are perfect reflectors, the zero order TM mode is unattenuated at all frequencies, and becomes what is sometimes called a TEM mode.

Now suppose that one surface of the wave-guide is doubly refracting. Then there are no purely TM or TE modes. The modes are, however, of two types. Those of one type are very similar to TM modes, but the magnetic field has a small component in the direction of propagation, and modes of the other type are similar to TE modes, but the electric field has a small component in the direction of propagation. We shall call these 'quasi-TM' and 'quasi-TE' modes, respectively. It is now possible for modes of both these types to be excited by a vertical electric dipole, although we shall show (§ 4.6) that the amplitudes of the quasi-TE modes are small.

Since quasi-TE modes may play some part in the propagation when the ionosphere is not isotropic, it is useful to discuss the propagation of TE-modes in the simple case where the earth and ionosphere are perfect reflectors. In such a guide the TE-modes could be excited by a vertical magnetic dipole, and we can discuss the excitation of the modes, using a point diffraction grating formed by the source and its images. Consider,

first, the image of the source formed by the ground. It is a vertical magnetic dipole of equal amplitude but opposite sign to the source. The source and its image may be considered to be a single element of the grating. A succession of images of this pair of dipoles is formed by the two surfaces of the wave-guide, and these form the remaining elements of the grating. Now a single element cannot radiate in a direction perpendicular to the grating, because the radiations from the two component dipoles in each element cancel. Hence the grating can have no spectrum with its wave normal parallel to the surfaces of the wave-guide. The spectra, or wave-guide modes, of lowest order occur where $2h \cos i = \pm\lambda$, and in general modes of higher order occur where $2h \cos i = \pm n\lambda$, where n is an integer other than zero.

The values of $\cos i$ for the various modes of both kinds are given by the positions of the poles of an integrand (eqn. (19), § 3.2, eqn. (30), § 3.3).

§ 3. Derivation of the Mathematical Expressions

3.1. *The Total Field of a Vertical Electric Dipole*

We now derive an expression for the field in the wave-guide due to a vertical electric dipole source. Cylindrical polar co-ordinates (ρ, ϕ, z) are used, and the source is at $(0, 0, H)$. The component of the field contributed by direct radiation from the dipole is given by (13). The space part of this expression may be written:

$$\{\exp(-jkr)\}/r = \tfrac{1}{2}jk \int_{-\infty}^{\infty} \mathrm{Hi}_0(kS\rho) \exp\{\mp jkC(z-H)\}\, dC, \quad \ldots \quad (15)$$

$$-\text{sign if } z > H,$$

$$+\text{sign if } z < H,$$

where $S = \sin i$, $C = \cos i$. (See, for example, Ott (1942), Sommerfeld (1935).) The integrand has branch points at $C = \pm 1$. The integral is a contour integral, and the contour must be indented above the real axis at $C = +1$, and below it at $C = -1$. The integrand is the equivalent of a set of plane waves distributed symmetrically in azimuth, and with their wave normals all at an angle i to the vertical. The field due to the original dipole is therefore given by

$$\mathbf{\Pi}_0 = \tfrac{1}{2}jk \int_{-\infty}^{\infty} \mathrm{Hi}_0(kS\rho) \exp\{\mp jkC(z-H)\}\mathbf{P}\, dC \quad \ldots \quad (16)$$

where $\mathbf{P}$ is given by (14).

As the evaluation of the total field for the series of image dipoles is complicated, we deal first with the case where the upper medium is isotropic, so that the reflection coefficient matrix $\mathbf{R}_2$ reduces to a pure number R_2 (a function of C), and $\mathbf{R}_g$ is unity. It may be helpful to visualize R_2 as a Fresnel reflection coefficient for a sharp boundary, although the analysis is not applied specifically to this case until § 4.

3.2. *The Residue Series—Isotropic Medium*

There are four series of images at heights $2mh+H$, $2mh-H$, $-2mh+H$, and $-2mh-H$, where m is a positive integer. The fields of these are given by

$$\mathbf{\Pi}_{r,m}=\tfrac{1}{2}jk\int_{-\infty}^{\infty}\mathrm{Hi}_0(kS\rho)\mathbf{P}f_{r,m}(C)\,.\,dC, \quad (r=1,\,2,\,3,\,4) \qquad . \quad (17)$$

where

$$f_{1,m}=[\exp\{jkC(z-H-2mh)\}]R_2^m, \quad (m=0,\,1,\,2,\,\ldots) \qquad (18a)$$

$$f_{2,m}=[\exp\{jkC(z+H-2mh)\}]R_2^m, \quad (m=1,\,2,\,\ldots) \qquad (18b)$$

$$f_{3,m}=[\exp\{-jkC(z-H+2mh)\}]R_2^m, \quad (m=1,\,2,\,\ldots) \qquad (18c)$$

$$f_{4,m}=[\exp\{-jkC(z+H+2mh)\}]R_2^m. \quad (m=0,\,1,\,2,\,\ldots) \qquad (18d)$$

The total field is given by

$$\mathbf{\Pi}=\tfrac{1}{2}jk\int_{-\infty}^{\infty}\mathrm{Hi}_0(kS\rho)\,.\,\mathbf{P}f(C)\,.\,dC, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where

$$f(C)=\sum_0^\infty(f_{1,m}+f_{4,m})+\sum_1^\infty(f_{2,m}+f_{3,m}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The series (20) are geometrical progressions, but before they can be summed we must examine their convergence. With the sign convention used in Sommerfeld's expression (15), the reflection coefficient R_2 reduces to the familiar reflection coefficient when C is real and positive. When C is negative, R_2 is in general greater than unity. For convergence of the series it is necessary that

$$|R_2\exp(-2j\alpha)|<1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where α stands for hkC. This holds on the positive real axis, but not on the negative real axis. We may, however, deform the contour in the integral (19) so that (21) holds everywhere on it. The original contour is shown in fig. 1 *a*. Let QQ′ be the line whose equation is

$$|R_2\exp(-2j\alpha)|=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

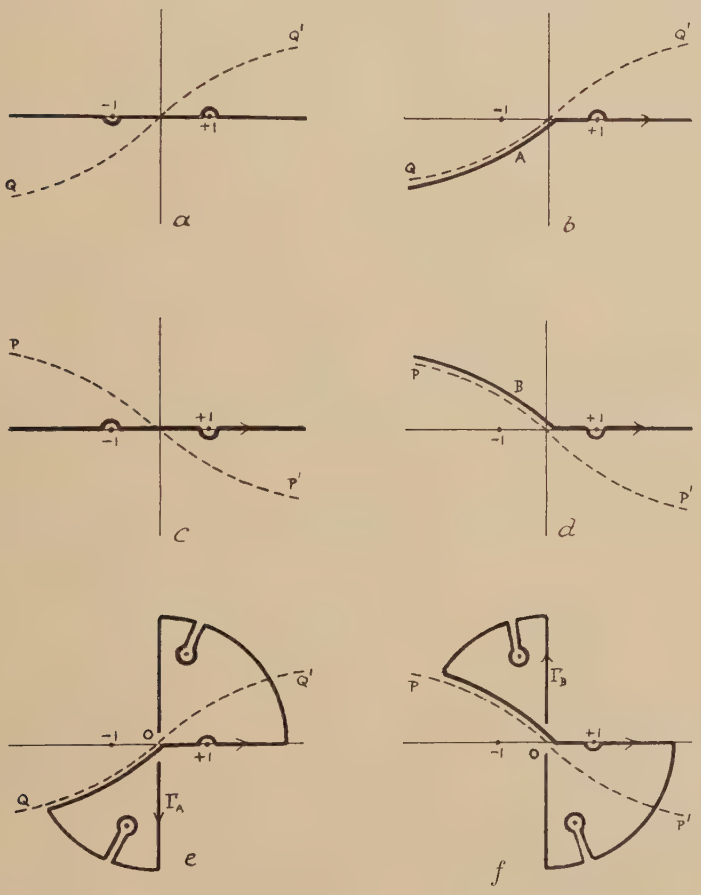
It can be shown that in the third quadrant the branch points of R_2 lie below the line QQ′, and there are no poles. The integrand vanishes at infinity, in the first and third quadrants, due to the term $\mathrm{Hi}_0(kS\rho)$. Hence the contour may be deformed to the position A, fig. 1 *b*, without crossing any singularity of the integrand, and (21) will then be valid. This applies for all the series (18 *a* to *d*). When this deformation is made, we may sum the series as geometrical progressions, and we have from (20)

$$f(C)=2\exp(-jkCH)\,.\,\cos(kCz)[1-R_2\exp(-2j\alpha)]^{-1}$$
$$+2\exp(jkCH)\,.\,\cos(kCz)R_2\exp(-2j\alpha)\,[1-R_2\exp(-2j\alpha)]^{-1}. \quad (23)$$

Suppose that in the integral (19) we replace C and S by their complex conjugates, C^* and S^*: this is not the same as replacing the whole integrand by its complex conjugate, since we do not replace j by $-j$,

except where it occurs in the variables C and S. The integrand is therefore replaced by a new function, and in particular it has different singularities. The values of C and S on the real axis are, however, unaltered by the change, and provided that we now use the contour of fig. 1 *c*, the value of the integral is the same as before. To make the series (18 *a* to *d*) convergent, we now deform the contour to the position B as shown in fig. 1 *d*, and the series can then be summed as geometrical progressions as before.

Fig. 1



Contours for integration in the complex C-plane.

We can therefore write (19) thus:

$$\mathbf{\Pi}=\tfrac{1}{4}jk\int_{\mathbf{A}}\mathrm{Hi}_0(kS\rho)\,.\,\mathbf{P}\,.\,f(C)\,.\,dD+\tfrac{1}{4}jk\int_{\mathbf{B}}\mathrm{Hi}_0(kS^*\rho)\,.\,\mathbf{P}\,.\,f(C^*)\,.\,dC^*, \quad . \; . \; . \quad (24)$$

where $f(C)$ is given by (23). $f(C)$ and $f(C^*)$ vanish everywhere at infinity.

Let the values of S, C, α at the poles of $f(C)$ be S_n, C_n, α_n. Then $f(C)$ has poles where

$$2\alpha_n + j \log_e R_2(C_n) = 2\pi n \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

(n is an integer), and there is a similar equation for the poles of $f(C^*)$.

Now suppose that the contours **A** and **B** are closed as in figs. 1 *e* and 1 *f*, into contours Γ_A and Γ_B. When the radius of the curved parts of these contours becomes indefinitely great, the contributions to the integrals from these curved parts vanish because of the term $\mathrm{Hi}_0(kS\rho)$. The contributions to the two integrals from the imaginary axis are equal in magnitude and opposite in sign since there S and S^* are purely real, and equal. Branch points of R_2 (and therefore of $f(C)$) are excluded from the contours by cuts as shown (figs. 1 *e*, and *f*). (There may be more than the two branch points shown.) The contributions to the integrals from these cuts are associated with waves of a type discussed by Sommerfeld (1909) (the surface wave), and by Ott (1942) (the ' head ' wave). For the particular cases discussed in § 4, the branch points occur where the imaginary part of S is large. The associated waves would therefore be heavily attenuated, and would make a negligible contribution to the signal received at a distance. They are not wave-guide modes, and we shall not discuss them. We shall evaluate only the contribution to the integral from poles, and we use the symbol Π_p to denote this contribution. Previous authors (Bremmer 1949, Rydbeck 1944) have called this contribution the ' residue series '.

The poles of $f(C)$ in Γ_A are on the line OQ′, and occur for positive (or zero) values of n in (25). Poles of $f(C^*)$ in Γ_B are on OP′ (figs. 1 *d*, 1 *f*) and occur for negative values of n. The residue of $f(C)$ at the pole C_n is

$$4\{\cos(kzC_n)\cos(kHC_n)\}/\{jhk - R_2'/2R_2\}, \quad . \quad . \quad . \quad . \quad (26)$$

with a corresponding expression for poles of $f(C^*)$. A dash, ′, denotes differentiation with respect to C. Hence

$$\Pi_p = (2\pi j/h)\Big[\sum_{n=0}^{\infty} \mathrm{Hi}_0(kS_n\rho)\cos(kzC_n)\cos(kHC_n)\{1 + jR_2'/(2hkR_2)\}^{-1} + \sum_{n=-1}^{-\infty} \mathrm{Hi}_0(kS_{-n}{}^*\rho)\cos(kzC_n{}^*)\cos(kHC_n{}^*)\{1 + jR_2'/(2hkR_2)\}^{-1}\Big]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

The term $R_2'/(2hkR_2)$ is small except at very low frequencies. It was neglected in Paper I, and the resulting equation was quoted in Paper I, eqn. (8). The terms of the series (27) refer to the various wave-guide modes. The characteristics of the modes are determined by eqn. (25), which gives the positions of the poles. This is equivalent to Eckersley's phase integral relation, and was discussed in Paper I.

3.3. *The Residue Series—Non-Isotropic Medium*

Analysis similar to that of the last section can be used where eqns. (17) to (20) are in matrix form. We shall now assume that the transmitter is on the ground, so that $H=0$. Instead of (23), we then have

$$f(C)=\{\exp(+jkCz)+\mathbf{R}_g\exp(-jkCz)\}\exp(2j\alpha)\{\exp(2j\alpha)-\mathbf{R}_2\}^{-1}$$
$$+\{\exp(-jkCz)+\mathbf{R}_g^{-1}\exp(jkCz)\}\{\exp(2j\alpha)-\mathbf{R}_2\}^{-1}\mathbf{R}_2. \quad (28)$$

This includes the reciprocal matrix

$$\{\exp(2j\alpha)-\mathbf{R}_2\}^{-1}, \quad (29)$$

whose terms can be evaluated. (28) then becomes

$$f(C)=\{\exp(+jkCz)+\mathbf{R}_g\exp(-jkCz)\}(1/\Delta)\mathbf{M}\exp(2j\alpha), \quad (30)$$

where

$$\Delta=\begin{vmatrix}\exp(2j\alpha)-{}_{\parallel}R_{\parallel}, & {}_{\perp}R_{\parallel}\\ -{}_{\parallel}R_{\perp}, & \exp(2j\alpha)+{}_{\perp}R_{\perp}\end{vmatrix}. \quad (31)$$

$$\mathbf{M}=\begin{pmatrix}\exp(2j\alpha)+{}_{\perp}R_{\perp}, & -{}_{\perp}R_{\parallel}\\ {}_{\parallel}R_{\perp}, & \exp(2j\alpha)-{}_{\parallel}R_{\parallel}\end{pmatrix}. \quad (32)$$

(30) has poles where

$$\Delta=0. \quad (33)$$

There are an infinite number of these, and as before, the position of any one pole depends upon an integer n. The residue at the pole at $C=C_n$ is

$$\{\exp(-jkC_nz)+\mathbf{R}_g\exp(jkC_nz)\}\{1/\Delta'(C_n)\}\mathbf{M}(C_n)\mathbf{R}_2(C_n). \quad (34)$$

To integrate eqn. (20) we close the contours as before (figs. 1 *e* and *f*). The algebra involved in evaluating the matrix products is tedious, and we give below only the final result for the pole of order n. If n is negative, S_n, C_n, must be replaced by S_n^*, C_n^*.

$$\Pi_{\parallel n}=-\pi k\mathrm{Hi}_0(kS_n\rho)\cos(kC_nz)[\exp(2j\alpha_n)+{}_{\perp}R_{\perp}]$$
$$\times\exp(2j\alpha_n)P_{\parallel}/\Delta'(C_n), \quad (35)$$

$$\Pi_{\perp n}=j\pi k\mathrm{Hi}_0(kS_n\rho)\sin(kC_nz)\{\exp(2j\alpha_n)\}_{\parallel}R_{\perp}P_{\parallel}/\Delta'(C_n) \quad (36)$$

where

$$\Delta'(C_n)=\{\exp(2j\alpha_n)+{}_{\perp}R_{\perp}\}\{2jhk\exp(2j\alpha_n)-{}_{\parallel}R_{\parallel}'\}$$
$$+\{\exp(2j\alpha_n)-{}_{\parallel}R_{\parallel}\}\{2jhk\exp(2j\alpha_n)+{}_{\perp}R_{\parallel}{}_{\parallel}R_{\perp}'\}+{}_{\parallel}R_{\perp}{}_{\perp}R_{\parallel}'. \quad (37)$$

It is easily checked that (35) reduces to one term of (27) for the isotropic case.

From eqns. (35) and (36) we may deduce the ratio of the horizontal components of the magnetic field in, and perpendicular to the direction of propagation, for the wave-guide mode of order n. From eqns. (9) and (11) this ratio is equal to

$$-c\{\partial\Pi_{\perp}/\partial z\}/\{\partial\Pi_{\parallel}/\partial t\}, \quad (38)$$

and using (35) and (36) it becomes

$$\{C_n{}_{\parallel}R_{\perp}\}/\{{}_{\perp}R_{\perp}+\exp(2jhkC_n)\}. \quad (39)$$

3.4. *The Curvature of the Earth*

So far we have assumed that the earth is flat, and that there is free space between the walls of the wave-guide. The curvature of the earth can be allowed for (e.g., Booker and Walkinshaw 1946) by considering the wave-guide to be filled with a fictitious medium having a vertical gradient of refractive index μ, given by

$$\mu=1+K\zeta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

where ζ is measured vertically upwards from the base of the top wall of the guide, and is therefore negative within the guide, and K is the curvature of the earth. If $Kh \ll 1$, we may incorporate this modification into the theory by using the phase integral method. We shall neglect the effect of the curvature on the ionized medium which forms the upper boundary of the guide, so that we assume that it remains homogeneous. Since the fictitious medium in the guide is not homogeneous, the component waves of any one mode now have curved wave-fronts. Let θ be the inclination to the vertical of the wave-normal of a component wave at a height $z=h+\zeta$. Let $q=\mu\cos\theta$. If $\theta=i$ at the surface of the ionosphere, then for any height in the guide

$$q^2=\mu^2-\sin^2 i \doteqdot 2K\zeta+\cos^2 i, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

since K is small.

$$\therefore q \doteqdot \cos i+K\zeta/\cos i. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

The phase integral relation is

$$\exp\{2j\int_{-h}^{0} kq\,.\,d\zeta\}=R_2\,.\,\exp(2\pi jn). \quad . \quad . \quad . \quad . \quad (43)$$

If $K=0$, the left hand side becomes $\exp(2jhkC)$ and the equation reduces to (22). If $K\neq 0$, we have for the left hand side of (43), using (41):

$$\exp\{2jhkC-jKkh^2/C\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

where $C=\cos i$.

Hence the curvature of the earth may be allowed for by replacing $\exp(2jhkC)$ by (44) in the formulae. This is an approximate method, but is accurate as long as the second term of (44) is very much less than the first, that is, as long as

$$|C| \geqslant \sqrt{(Kh/2)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

This was true for the numerical results discussed in §§ 4.4, 4.5 and 4.6.

§4. Application to the Ionosphere

4.1. *The Reflection Coefficients of the Ionosphere*

The preceding theory is applicable to any model of the ionosphere, for which the reflection coefficients are known analytic functions of the angle of incidence, regarded as a complex variable.

Before we can proceed further, it is necessary to adopt some specific model for the ionosphere. Theories which take account of the rate at which ionospheric characteristics change with height (e.g. Wilkes 1947, Heading and Whipple 1952), do not give sufficiently simple analytic functions for this purpose.

In Paper I the ionosphere was assumed to be a homogeneous sharply bounded medium containing N electrons per c.c., subject to collisions at an average rate ν per second, but the earth's magnetic field was neglected. A discussion has been given elsewhere (Budden 1951 b) of the reflection of very low frequency radio waves when the ionosphere is assumed to be homogeneous and sharply bounded, with the effect of the earth's magnetic field included. In a full treatment of this problem the refractive index μ is given by the Appleton–Hartree magneto-ionic formula (Appleton 1932). It was shown, however, that fair agreement with the experimental results at very low frequencies could be obtained using the quasi-longitudinal approximation to the magneto-ionic formula (Booker 1935) and that this approximation becomes more accurate as the frequency decreases. We are therefore encouraged to use it in discussing the propagation of radio atmospherics. We shall now apply it to the results of the preceding sections.

Using this theory, we find (Budden, *loc. cit.*) that

$$\mu^2=1-(jp_r/p)\exp(\pm j\tau), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

where p_r and τ are functions of the characteristics of the ionosphere, and do not include the wave frequency. It can be shown that approximately

$$p_r=p_0^2/\sqrt{(\nu^2+p_L^2)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

$$\tan\tau=p_L/\nu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

where

$p_0^2=4\pi Ne^2/\epsilon_0 m$.

ν=collision frequency.

$p_L=\mu_0 eH_L/m$.

e, m=charge and mass of the electron.

N=number of electrons per unit volume.

H_L=component of earth's magnetic field in direction of propagation.

ϵ_0, μ_0=electric and magnetic permittivities of free space. (The symbol μ_o is used later to mean the refractive index of the ordinary wave, and must not be confused with the μ_0 used here.)

A plane wave incident on the ionosphere at angle of incidence i gives rise to a reflected wave below the ionosphere, and to both ordinary and extraordinary waves in the ionosphere. For each wave $E_{\parallel}$ denotes the component of **E** in the plane of incidence and perpendicular to the wave

normal, and $E_\perp$ denotes the component perpendicular to the plane of incidence. Then for upgoing waves in the northern hemisphere we have for the ordinary wave

$$\mu_o^2=1-(jp_r/p)\exp(j\tau), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

$$E_{\perp o}/E_{\parallel o}=-j, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

and for the extraordinary wave

$$\mu_e^2=1-(jp_r/p)\exp(-j\tau), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

$$E_{\perp e}/E_{\parallel e}=+j. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

The wave normals of these two waves are assumed to make angles θ_o and θ_e respectively to the vertical. The angles i, θ_o and θ_e are related by Snell's law.

$$\sin i=\mu_o \sin\theta_o=\mu_e \sin\theta_e. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

The expressions (49) to (53) were used, together with the boundary conditions which apply at the interface between the ionized medium and the free space below it (Budden, *loc. cit.*), to find the components of **R**. The results apply both to $E_\parallel$ and $E_\perp$ and to the components $\Pi_\parallel$ $\Pi_\perp$ of the Hertz vectors, and are as follows:

$${}_\parallel R_\parallel=\{(\mu_o\mu_e-1)(\cos\theta_o+\cos\theta_e)\cos i+(\mu_o+\mu_e)(\cos^2 i-\cos\theta_o\cos\theta_e)\}/D,$$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

$${}_\parallel R_\perp=2j\cos i(\mu_o\cos\theta_o-\mu_e\cos\theta_e)/D, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

$${}_\perp R_\parallel=2j\cos i(\mu_o\cos\theta_e-\mu_e\cos\theta_o)/D, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

$${}_\perp R_\perp=\{-(\mu_o\mu_e-1)(\cos\theta_o+\cos\theta_e)\cos i+(\mu_o+\mu_e)(\cos^2 i-\cos\theta_o\cos\theta_e)\}/D,$$

$$. \quad . \quad . \quad . \quad . \quad (57)$$

where

$$D=(\mu_o\mu_e+1)(\cos\theta_o+\cos\theta_e)\cos i+(\mu_o+\mu_e)(\cos^2 i+\cos\theta_o\cos\theta_e). \quad (58)$$

4.2. *Method of Obtaining Numerical Solutions*

In the following sections we give some results of calculations made using the preceding formulae. To find the characteristics of any one mode, it is necessary to find the positions of the poles of the integrand in eqn. (20) by solving eqns. (31), (33) using the expressions (46) to (58). Let

$$hkC=\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

Then eqn. (33) reduces to the following quadratic equation for $\tan\alpha$:

$$\mu_o\mu_e\tan^2\alpha-j\tan\alpha(\mu_o+\mu_e)(C^2+\cos\theta_o\cos\theta_e)/\{C(\cos\theta_o+\cos\theta_e)\}-1=0.$$

$$. \quad . \quad . \quad . \quad . \quad . \quad (60)$$

It can be shown that one root of this quadratic gives the quasi-TM modes, and the other gives the quasi-TE modes. Suitable values of the parameters p_r, τ, and h were selected as described in the next section.

The type and order of the mode for which solutions were required was also chosen. Then the equation was solved for each of a series of values of the angular frequency p. The method of solution was as follows. A trial value of C was assumed, and from it the values of $\cos\theta_o$ and $\cos\theta_e$ were calculated, using (53). The quadratic was then solved for $\tan\alpha$ from which α was calculated. Then (59) gave a revised value of C which was used to suggest a new trial value of C. The process was repeated until a self-consistent solution was obtained. The actual calculations were made using the EDSAC calculating machine in the Mathematical Laboratory of the University of Cambridge.

In a few cases the curvature of the earth was allowed for, by using the equation

$$\alpha = hkC - Kkh^2/2C \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

instead of eqn. (59). Equation (61) is derived from (44). The effect of the earth's curvature was found to be very small, and in most of the calculations the earth was assumed to be flat.

It is also of interest to calculate, for any one mode, the ratio of the two horizontal components of the magnetic field of the wave, namely, the component in the direction of propagation, and the component at right angles to it. For an isotropic reflecting medium, this ratio is zero for a TM mode, and infinite for a TE mode. In general its value is given by (38), for which an algebraic expression was obtained. This was very complicated and will not be given here. We shall call this ratio the ' polarization factor '. The results of the calculations are given in § 4.6.

Finally we wish to known how much of a given mode is excited by a vertical electric dipole. For this purpose, a quantity was calculated which is proportional to the transverse horizontal component of the magnetic field of the wave, when the attenuation term is omitted. The expression used was derived from eqn. (35) using eqn. (9). It will be called the ' excitation factor '. The results of the calculations are discussed in § 4.7.

4.3. *The Values of the Parameters*

The parameters p, p_r, h, and τ occur only in the combinations p/p_r, $h\,.\,p_r$, and τ. Solutions were obtained for fixed values of $h\,.\,p_r$ and τ, and for a series of values of p/p_r (equivalent to a series of different frequencies, p). The frequency scales in the figures in this paper are chosen on the assumption that $h=80$ km.

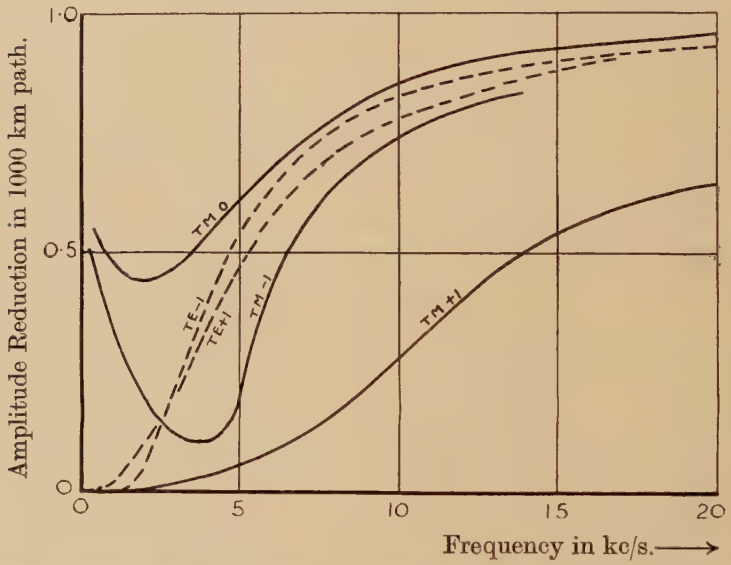
In a previous paper (Budden 1951 b), the results were given of some calculations of the reflection coefficients (54) and (55) for real angles of incidence, and these were compared with experimental measurements for a frequency of 16 kc/s (Bracewell *et al.* 1951). The values of the parameters p_r and τ which gave best agreement with the experimental results were $p_r=2\times10^5\ \mathrm{sec}^{-1}$, $\tau=60°$. Most of the following results therefore apply to these values.

In Paper I, some results were given of calculations in which the earth's magnetic field was neglected, that is $\tau=0$, for several values of p_r. All the modes in this case were pure TM modes. Some results for $\tau=0$, $p_r=2\times10^5$ sec^{-1} are also given in this paper for comparison with the results when $\tau=60°$. They refer to TM modes of order -1, 0, and $+1$, and to TE modes of order -1 and $+1$. These TE modes cannot be excited by a vertical electric dipole, but it is interesting to compare them with the quasi-TE modes which occur if the top reflector is doubly refracting, and which can be excited by a vertical electric dipole.

4.4. *Numerical Results for Attenuation in Wave-Guide Modes*

In fig. 2, curves of attenuation versus frequency are shown for $p_r=2\times10^5$ sec^{-1}, $\tau=0°$, $h=80$ km. The full curves are for TM modes,

Fig. 2



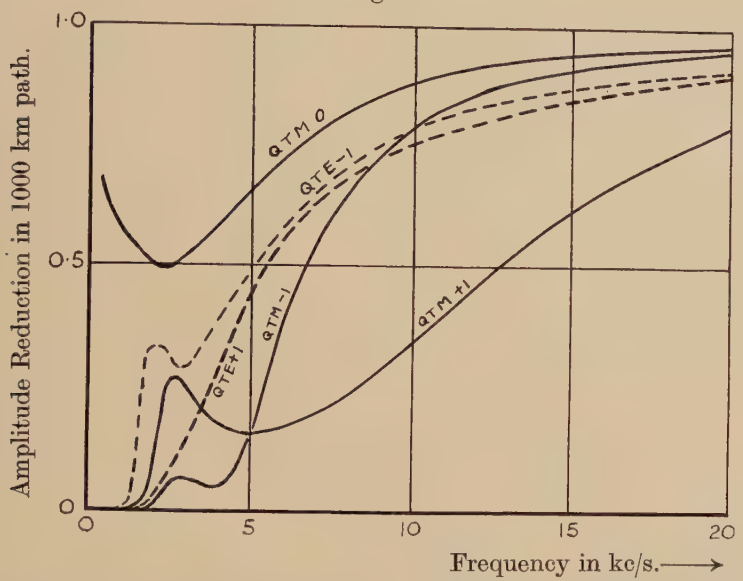
Fractional reduction in received e.m.f. for 1000 km path in modes of various order. $h=80$ km, $p_r=2\times10^5$ sec^{-1}, $\tau=0°$.

and the dotted curves for TE modes. In fig. 3, the corresponding curves are shown for $p_r=2\times10^5$ sec^{-1}, $\tau=60°$, $h=80$ km: the full curves are for quasi-TM modes, and the dotted curves for quasi-TE modes.

It is clear that the main features of the curves are the same both for $\tau=0°$, and $\tau=60°$. In particular, the quasi-TM mode of order zero is attenuated most heavily in a range of frequencies from about 1 kc/s to 8 kc/s. This frequency range is determined mainly by the height of reflection, h, and only to a smaller extent by the characteristics of the top reflector. This is illustrated in fig. 4, which shows the attenuation in the zero order mode for various values of p_r when $h=80$ km and $\tau=60°$.

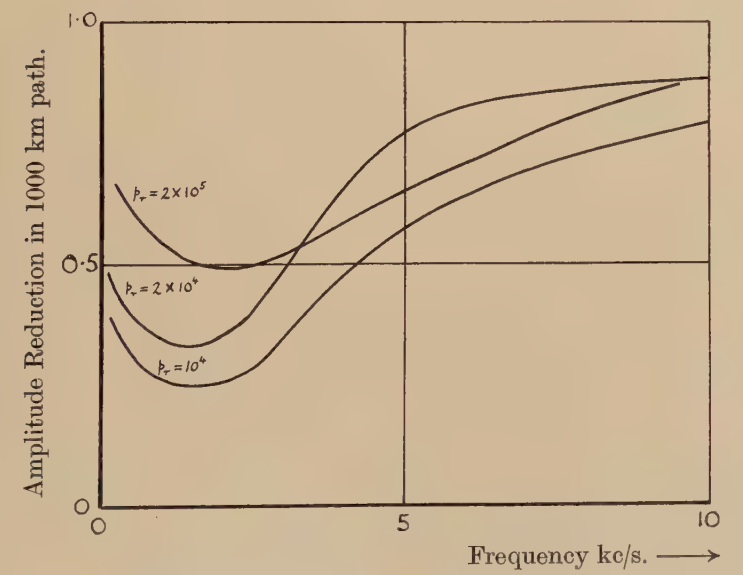
It was shown in Paper I, and is apparent from fig. 2 of this paper, that when $\tau=0$, the TM modes of orders 0, and -1, have very similar characteristics. This similarity disappears when $\tau=60°$. Then the quasi-TM mode

Fig. 3



Fractional reduction in received e.m.f. for 1000 km path in modes of various order. $h=80$ km, $p_r=2\times10^5$ sec^{-1}, $\tau=60°$.

Fig. 4

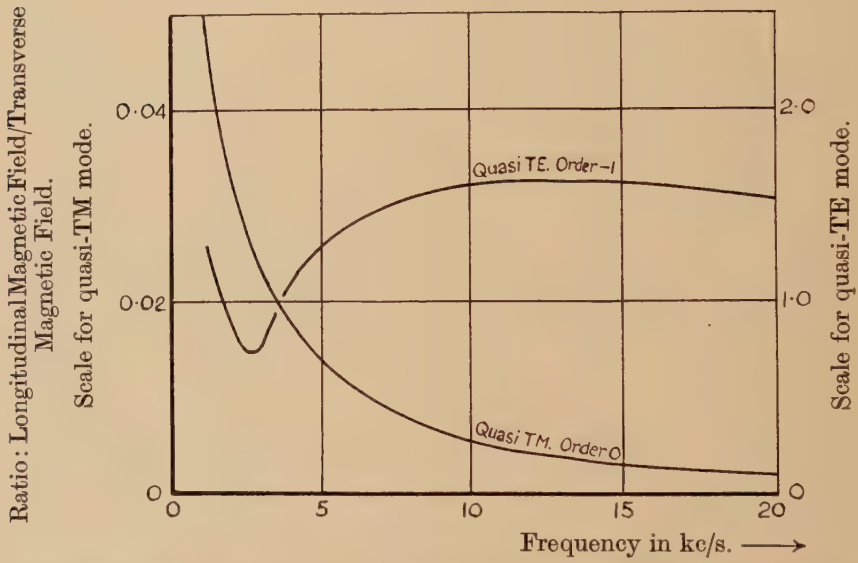


Fractional reduction in received e.m.f. for 1000 km path in the zero order mode, as a function of frequency. $\tau=60°$, $h=80$ km.

of order -1 is much more heavily attenuated, and its characteristics resemble more closely those of the quasi-TM mode of order $+1$.

The attenuation was also calculated for the quasi-TM mode of order zero when the earth's curvature was taken into account. It was found to be only very slightly greater (1·3% for a frequency of 5 kc/s), so that the effect of the curvature is very small. We should expect a larger attenuation because a wave which leaves the ground with a given angle of elevation is reflected from the ionosphere at an angle which is more nearly vertical if the earth is curved than if it is assumed to be flat. At nearly grazing incidence, the reflection coefficient decreases as the angle of incidence becomes more vertical (this is clearly shown in the curves given by Budden 1951 b). Thus we should expect heavier attenuation when the earth is

Fig. 5



Polarization factor as a function of frequency, for quasi-TM mode of order 0, and quasi-TE mode of order -1. $p_r=2\times10^5$ sec^{-1}, $\tau=60°$, $h=80$ km.

curved. The effect would become smaller if the angle of incidence were markedly different from grazing, and hence we should expect the earth's curvature to have a still smaller effect on modes of order other than zero.

4.5. *Numerical Results for Polarization Factor*

Polarization factors were calculated as described in § 4.2 for the quasi-TM mode of order 0, and the quasi-TE mode of order -1. The polarization factor is the ratio of the complex amplitudes of the horizontal components of the magnetic field in, and perpendicular to, the plane of incidence. The modulus of the polarization factor is plotted as a function

of frequency in fig. 5 for $p_r=2\times10^5$ sec^{-1}, $\tau=60°$, $h=80$ km. It is seen to be extremely small for the quasi-TM mode of order 0. It might be expected that the signal received from a distant low frequency transmitting station would consist predominantly of the quasi-TM mode of order zero, and hence we should expect that, if a rotatable loop receiving aerial were used, the bearing of the station would be sharp. This is known to be the case for signals from distant low frequency stations (see, for example, Round *et al.* 1925, Smith-Rose 1929).

For the quasi-TE mode, the polarization factor is seen to be much larger. Hence, if a signal consisting predominantly of a quasi-TE mode were received on a rotatable loop receiving aerial, the orientation of the loop for maximum or minimum signal would bear no simple relation to the direction of the source. We shall see, in the next section, however, that the amplitude of a quasi-TE mode excited by a vertical electric dipole, is very small compared with the amplitude of the quasi-TM mode of order zero. It is therefore probable that quasi-TE modes play little part in the propagation to great distances of signals from very low frequency transmitters.

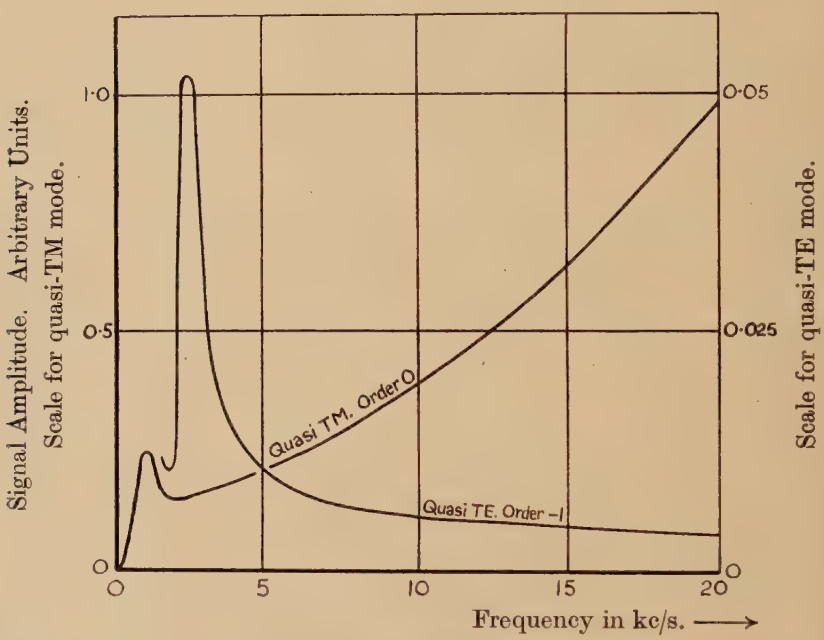
Many experiments are known in which radio atmospherics are received on a pair of crossed loop receiving aerials. For example, the atmospheric direction finding chain of the British Meteorological Office uses a number of receiving stations of this type (see Adcock and Clarke 1947), in which the configuration of the horizontal magnetic field in the received signal is displayed on the screen of a cathode ray oscillograph. It is well known that most atmospherics received on a system of this type give a response which is practically a straight line on the screen. Hence the magnetic field of the wave is almost entirely in one direction which is assumed to be perpendicular to the direction of propagation, and a good bearing determination is then possible on the source of the atmospheric. This suggests that quasi-TE modes, if present, have a very small amplitude. Occasionally, however, an atmospheric is received which gives a very elliptical or irregular response on the screen, and no accurate bearing measurement can be made. It is possible that this type of atmospheric contains an abnormally large proportion of quasi-TE modes. Since these would not normally be excited with large amplitudes by a vertical lightning flash, we must assume that the currents in the flash where they originate have a large horizontal component.

4.6. *Numerical Results for Excitation Factor*

Excitation factors were calculated as described in § 4.2 for the quasi-TM mode of order 0, and the quasi-TE mode of order -1. The excitation factor is a complex quantity whose modulus is proportional to the amplitude of the transverse horizontal component of the magnetic field in the wave at a given distance from the source, when the attenuation term is omitted. The modulus of the excitation factor is plotted as a function of frequency in fig. 6, for $p_r=2\times10^5$ sec^{-1}, $\tau=60°$, $h=80$ km.

The amplitude scale used for the quasi-TM mode of order 0 is 20 times the scale for the quasi-TE mode of order -1. From fig. 6 we see, for example, that for a frequency of 10 kc/s, the excited amplitude for the quasi-TM mode is 72 times the amplitude for the quasi-TE mode. For all frequencies above about 5 kc/s, the amplitude of the quasi-TE mode is very small compared with the amplitude of the quasi-TM mode. For lower frequencies the excitation factors become more nearly comparable, but in this range of frequency, the quasi-TE mode is heavily attenuated (see fig. 3). Hence, at a distance from a vertical dipole source, we should not expect quasi-TE modes to appear with appreciable amplitude.

Fig. 6



Excitation factor as a function of frequency, for quasi-TM mode of order 0, and quasi-TE mode of order -1. $p_r=2\times10^5$ sec^{-1}, $\tau=60°$, $h=80$ km.

4.7. *The Signal Received from an Atmospheric*

From the curves of excitation factor and attenuation as functions of frequency, it would be possible to calculate the signal received at a distance from a single atmospheric. This was done in Paper I, for the case when the earth's magnetic field was neglected. The method involves the evaluation of a Fourier integral. The impulse responses of this type have not been calculated when allowance is made for the earth's magnetic field, but since the attenuation curves are similar in form, whether or not the effect of the field is included, we should expect the results in the two cases to be very similar.

§ 5. Conclusions

The propagation of low frequency radio waves may be treated either in terms of successive reflections at the ionosphere, or in terms of wave guide modes. When the separation of the transmitter and receiver is large compared with the height of the ionosphere, it is simpler to use wave guide modes. The treatment given in Paper I, in which the earth's magnetic field was neglected, led to a qualitative explanation of some of the phenomena associated with the propagation of radio atmospherics to great distances.

The earth's magnetic field is important in the theory of the ionosphere, and it was therefore necessary to investigate its effect on the wave-guide modes. This has been done in the present paper, and for the model discussed it has been shown that there is little change in the general form of the results, so that the treatment given in Paper I is justified. When the earth's magnetic field is included, the modes are somewhat more complicated, and a method of calculating their polarization characteristics and degree of excitation has been given.

The modes may also be affected by the curvature of the earth, which was neglected in Paper I. The present paper shows that this was probably justified, for frequencies below about 10 kc/s. At higher frequencies, however, it is possible that the earth's curvature may have an appreciable effect on the modes.

In some recent experiments made by Dr. Weekes, the signal from a low frequency radio transmitter (frequency 16 kc/s), was measured for a range of distances extending up to 3650 km, from the transmitter. It is believed that the results of these measurements may be simply interpreted in terms of wave-guide modes.

Acknowledgments

This work was carried out at the Cavendish Laboratory as part of a programme of Radio Research. I am indebted to Dr. Wilkes, Director of the Mathematical Laboratory, University of Cambridge, and his staff, for permission to use the EDSAC calculating machine, and for able instruction in its use, and especially to Dr. R. A. Brooker who devised the programme for the EDSAC and gave invaluable help while the calculations were in progress. I am also indebted to Professor Hartree and Mr. Ratcliffe for valuable advice and discussion.

References

Adcock, F., and Clarke, C., 1947, *J. Instn. Elect. Engrs.*, **94,** III, 118.
Appleton, E. V., 1932, *J. Instn. Elect. Engrs.*, **71,** 642.
Booker, H. G., 1935, *Proc. Roy. Soc.* A, **150,** 267.
Booker, H. G., and Walkinshaw, W., 1946, *Report on Meteorological Factors in Radio-Wave Propagation* (London : Physical Society).
Bracewell, R. N., Budden, K. G., Ratcliffe, J. A., Straker, T. W., and Weekes, K., 1951, *Proc. Instn. Elect. Engrs.*, **98,** III, 221.
Bremmer, H., 1949, *Terrestrial Radio Waves* (Amsterdam : Elsevier Publishing Co.).

Budden, K. G., 1951 a, *Phil. Mag.*, **42,** 1 ; 1951 b, *Ibid.*, **42,** 833.
Eckersley, T. L., 1932, *Proc. Roy. Soc.*, **136,** 499.
Heading, J., and Whipple, R. T. P., 1952, *Phil. Trans. Roy. Soc.* A, **244,** 469.
Ott, H., 1942, *Ann. der Phys.*, **41,** 443.
Round, H. J., Eckersley, T. L., Tremellen, K., and Lunnon, F. C., 1925, *J. Instn. Elect. Engrs.*, **63,** 933.
Rydbeck, O., 1944, "On the Propagation of Radio Waves". *Trans. Chalmers Univ., Gothenberg, Sweden*, Nr. 34.
Smith-Rose, R. L., 1929, *Proc. Inst. Radio Engrs.*, **17,** 425.
Sommerfeld, A., 1909, *Ann. der Phys.*, **28,** 665 ; 1935. Article in ***Frank und v. Mises*** : ***Differential Gleichungen der Physik.***
Wilkes, M. V., 1947, *Proc. Roy. Soc.* A, **189,** 130.

CXIX. *The Mean Lifetime and Frequency of Production of Charged V-Particles*

By K. H. Barker, C. C. Butler, M. G. Sowerby and C. M. York
The Physical Laboratories, University of Manchester*

[Received August 19, 1952]

Abstract

Data for the determination of the mean lifetime of charged V-particles are discussed, following the procedure outlined by Wilson and Butler (1952). The tracks of twenty-seven $V^{\pm}$-particles have now been photographed on the Pic-du-Midi (2 867 m) by Armenteros *et al.* (1952) and by the authors in a circular cloud chamber, 28 cm in diameter. It is shown that, using these data, the chamber is too small for the lifetime to be determined, but a lower limit of $1{\cdot}0\times10^{-10}$ sec is obtained. Furthermore, the relation between the frequency of production of energetic $V^{\pm}$-particles in penetrating showers of average energy (10–30) Bev, and various assumed lifetimes is discussed. Using recent photographic emulsion data on the frequency of particles of mass about 1 200 m_e in the same type of penetrating shower an upper limit for the lifetime of 10^{-8} sec is obtained.

§ 1. Introduction

In a recent letter Wilson and Butler (1952) discussed several methods for measurement of the lifetime of unstable particles, such as charged V-particles. They concluded that the most direct procedure is to deduce a value from cloud-chamber photographs showing recognizable decays of $V^{\pm}$-particles. The decays of twenty-seven $V^{\pm}$-particles have been photographed on the Pic-du-Midi (2 867 m), using a circular cloud chamber 28 cm in diameter. The momenta and angles of apparent deflection for twenty-two of these decays were given by Armenteros *et al.* (1952) in table 1 of their paper. A further five examples have now been obtained by the authors ; this additional data is summarized in table 1.

Table 1. Measurements on Five Charged V-Particles†

(1) Event No.	(2) Sign of V	(3) Primary momentum (100 Mev/c)	(4) $I/I_{\min}$ of primary	(5) Angle ϕ (deg)	(6) Sign and momentum of secondary (100 Mev/c)	(7) Transverse momentum P_T (100 Mev/c)
23	+	—	1	13·5	$+11{\cdot}4\pm2{\cdot}3$	$2{\cdot}60\pm0{\cdot}50$
24	?	—	1	15–20	—	—
25	−	—	1	18	$-6{\cdot}0\pm1{\cdot}0$	$1{\cdot}86\pm0{\cdot}31$
26	−	—	1	10	-16 ± 5	$2{\cdot}70\pm0{\cdot}90$
27	+	—	1	19	$+8{\cdot}0\pm1{\cdot}6$	$2{\cdot}64\pm0{\cdot}53$

* Communicated by Professor P. M. S. Blackett, F.R.S.
† This table has the same form as table 1 of Armenteros *et al.* (1952).

§ 2. The Experimental Data

In order to determine the lifetime of $V^{\pm}$-particles it is necessary to know two quantities for each decay in addition to those given in table 1. Firstly, the length of the track produced by the unstable particle in the illuminated region of the chamber, and secondly, the distance the particle would have travelled in the chamber if it had not decayed must be measured. Wilson and Butler called these two lengths the ' path length ' and the ' potential path length ' respectively. Thus decay takes place after a path length, x, which is always shorter than the potential path length, x_0. Twenty-three pairs of values of x and x_0 are given in columns (2) and (3) of table 2. The tracks of the remaining four $V^{\pm}$-particles are either too short for any measurements to be made or they are situated near the edge of the chamber in regions where stereoscopic measurements cannot be made.

The values of x and x_0 have been determined by reprojecting the stereoscopic photographs. In most cases x is easily measured, the estimated errors are given in column (2) of table 2; only approximate values can be given for a few of the events. The charged V-particles 3 and 10 were produced in a lead plate across the centre of the chamber; x cannot be measured accurately for 3 because the decay occurred among a great many penetrating shower particles, the tracks of which prevent accurate reprojection of the decay. The lead plate was also in use when events 1, 2 and 8 were obtained. When either of the lengths x or x_0 crossed the plate its thickness has been subtracted from the measured values.

In order to measure x_0 the track of each unstable particle has to be extrapolated until it leaves the illuminated region of the chamber. There is little difficulty in finding where the charged V-particle entered the fully illuminated volume of the chamber, but there is often some uncertainty in the determination of the exit point, supposing that the particle had not decayed. The illuminated depth of the chamber is about 7 cm; it is important, however, to determined this depth for each photograph because the illumination conditions may vary during the operation of the cloud chamber. Most of the decays of $V^{\pm}$-particles are associated with showers of several particles, and the illuminated depth can be determined using these particles. If this procedure is not convenient, penetrating showers on the same piece of film are used. The estimated errors in the values of x_0 are given in column (3) of table 2; in several cases it is impossible to give reasonably accurate values of x_0.

Very few of the observed $V^{\pm}$-particle decays took place very close to the top of the cloud chamber. This is probably due to the difficulty of identifying decays which take place near the top of the chamber (or immediately underneath the lead plate) because of the confusion produced by the tracks of other particles in the penetrating showers. There is, therefore, a bias against selecting photographs showing decays with small x. Wilson and Butler pointed out this difficulty and suggested that the values of x and x_0 should be measured from fiducial surfaces, draw

Table 2. Data on the Decays of Twenty-three $V^{\pm}$-Particles

(1) Event No.	(2) x (cm)	(3) x_0 (cm)	(4) x' (cm)	(5) x_0' (cm)	(6) x'/x_0'	(7) p (100 Mev/c)	(8) ϕ (deg)	(9) P/M	(10) t (10^{-10} sec)	(11) T (10^{-10} sec)
1	~2	~6	—	—	—	~20	19	—	—	—
2	8·4±1·0	16·6±2·0	5·4	12·6	0·43±0·10	15±5	24·5	2·5	0·7	1·7
3	~2	~7	—	—	—	~2	10·3	0·3	—	—
6	6·6±0·5	16·1±1·0	3·6	12·1	0·30±0·05	11±4	12·5	1·8	0·7	2·2
7	3·2±0·5	25·2±2·0	0·2	21·2	0·01±0·25	21*	9·5	3·3	0·1	2·1
8	4·3±0·3	10·9±2·0	1·3	6·9	0·19±0·07	~2	68·5	0·3	0·2	8·0
9	5·8±1·0	19·9±2·0	2·8	15·9	0·18±0·07	8·0±2·6	15	1·3	0·7	4·1
10	1·1±0·1	6·6±1·0	—	—	—	7*	34	1·2	—	—
11	14·7±1·0	20·3±3·0	11·7	16·3	0·72±0·15	6·0±0·6	22	1·0	3·9	5·4
12	12·2±1·0	22·0±2·0	9·2	18·0	0·51±0·08	15±3	27·5	2·5	1·2	2·4
13	14·7±1·0	23·0±2·0	11·7	19·0	0·62±0·09	50±20	15	8·3	0·5	0·8
15	6·7±1·0	13·5±2·0	3·7	9·5	0·39±0·13	6·0±1·5	9·5	1·0	1·2	3·2
16	3·9±0·5	19·6±5·0	0·9	15·6	0·06±0·04	16*	25	2·6	0·1	2·0
17	6·8±0·5	13·9±4·0	3·8	9·9	0·38±0·16	17*	20·5	2·7	0·5	1·2
18	~12	~19	—	—	—	>40	7	~8	—	—
19	9·8±0·5	18·5±2·0	6·8	14·5	0·47±0·07	>40	21·5	~8	0·3	0·6
20	~2	—	—	—	—	13*	34	2·1	—	—
21	14·8±0·5	21·8±1·0	11·8	17·8	0·66±0·05	>40	5	~8	0·5	0·7
22	13·0±0·5	18·3±1·0	10·0	14·3	0·70±0·06	4·1±0·4	60	0·7	5·0	6·8
23	7·2±0·5	21·0±1·0	4·2	17·0	0·25±0·03	28*	13·5	4·6	0·3	1·2
25	8·0±0·5	15·4±2·5	5·0	11·4	0·44±0·10	18*	18	2·5	0·7	1·5
26	8·8±0·5	19·2±1·0	5·8	15·2	0·38±0·04	40*	10	6·6	0·3	0·8
27	3·1±0·5	22·2±2·0	0·1	18·2	0·01±0·05	20*	19	3·3	0·01	1·8

Column (7) : An asterisk indicates that the particular P is estimated.
Column (8) : ϕ is the angle of deflection of the V-track.

a predetermined distance within the chamber. For the present data it is convenient to measure x and x_0 from the surfaces of the well-illuminated volume and then to reduce the values of x and x_0 (x_0 must also be reduced since it is measured from the beginning of the path length) by a constant amount. This length is sufficient for a decay to be identified and also for the direction of the charged V-particle to be found with reasonable precision. The values of x have all been reduced by 3 cm; this figure is somewhat arbitrary and is discussed again in § 3. Four decays with values of x less than 3 cm are rejected from the analysis. The corrected values of x, namely x', are given in column (3) of table 2.

Similarly, it is very difficult to detect the decay of a charged V-particle very close to the end of its potential path length. To avoid the possibility of a bias against observing this type of event all the values of x_0 are reduced by 1 cm. This distance should be long enough for the apparent deflection of a track to be recognized, and thus the event identified. The corrected values of x_0, given as x_0' in column (5) of table 2, are obtained by reducing the values of x_0 by 4 cm.

To measure the lifetime of the unstable particles it is necessary to convert the distances x' and x_0' into the proper times of flight. The momenta and masses of the unstable particles must be known before this transformation can be made. Provisionally it is assumed that the charged V-particles are all of the same kind and have a mass of approximately 1200 m_e; this assumption is discussed in § 3.

The momenta, P, of nine of the $V^{\pm}$-particles have been measured; the values are given in column (7) of table 1. In addition, three approximate lower limits are given for three high-energy events. Photographs 3 and 8 show heavily-ionizing primary particles, and the momenta given for these have been determined from the estimated ionizations and the assumed mass. Rough estimates of the momenta of the remaining nine events, which all have short primary tracks, can be made because the average value of the ratio of the primary to the secondary momentum, for the events with measured primaries, is very approximately 2·5. This rule was used when making the estimates labelled with an asterisk in column (7) of table 2.

The values of the dilatation factor, P/M, are given in column (9) of table 2. P is the momentum and M is the rest mass, in energy units, of each charged V-particle. The proper times of flight, t, and the potential traversal times, T, are given in columns (10) and (11) of table 2. These values are obtained by dividing x' and x_0' by Pc/M, where c is the velocity of light.

§ 3. The Mean Lifetime of Charged V-Particles

In principle, the mean lifetime of the charged V-particles can now be calculated from the pairs of times t and T. These times have been measured assuming that all the $V^{\pm}$-particles have a mass of 1200 m_e.

Armenteros *et al.* (1952) concluded that $V^{\pm}$-particles are probably similar to the S-particles found by Bridge and Annis (1951) and Annis *et al.* (1952) and also to the κ-particles found by O'Ceallaigh (1951). The S- and κ-particles have a mass of about $1\,200\,m_e$ and decay at rest into a single charged meson. Recent work at Bristol, reported by Powell (1952), shows that the $1\,200\,m_e$ mass mesons observed in photographic emulsions are not all the same. There now appear to be at least two types, with masses of about $1\,100\,m_e$ and $1\,500\,m_e$. It is probable that the $V^{\pm}$-particles are also not homogeneous, and therefore it is desirable to classify the particles responsible for the observed decays before proceeding to the lifetime determinations. Unfortunately, at present, the decays-in-flight observed in the cloud chamber cannot be classified unambiguously and, therefore, the data can only be analysed on the assumption that all the $V^{\pm}$-particles have an average mass of about $1\,200\,m_e$. Owing to the small number of decays available, no evidence for the existence of two types of $V^{\pm}$-particle can be obtained from the lifetime measurements. Unfortunately there is no possibility of measuring the separate lifetimes of different types if they occur in about equal numbers, even if they have appreciably different lifetimes.

A preliminary examination of the usefulness of the data given in table 2 can be made by using the ratios x'/x_0' given in column (6). It can easily be shown that if the distribution of a large number of values of x'/x_0' is almost uniform from 0 to 1, with an average value close to 0·5, then the lifetime could be much longer than the mean traversal time, and no significant upper limit can be obtained. In fact the average value of x'/x_0' for eighteen decays is 0·38 and the mean value of T is $2{\cdot}6\times10^{-10}$ sec. There are two sources of error in this determination of the average x'/x_0'. Firstly, the error due to the individual errors of measurement of x' and x_0' is 0·03; this approximate value is calculated on the assumption that the errors in columns (2) and (3) of table 2 are independent. Secondly, there is a larger statistical error due to the fact that only a small number of ratios is available; an estimate of this error is 0·07. The average ratio, therefore, is not significantly smaller than 0·5, consequently the lifetime could be very much longer than $2{\cdot}6\times10^{-10}$ sec, and no upper limit to the value can be calculated. In order to obtain a lifetime using a sample of about 20 events, the average ratio x'/x_0' needs to be less than about 0·25.

Owing to the observed form of the momentum spectrum of the $V^{\pm}$-particles and to the geometry of the cloud chamber the values of T are spread from $0{\cdot}6\times10^{-10}$ sec to $8{\cdot}0\times10^{-10}$ sec. It is obvious that the most important decays are those with the longer potential traversal times. Photographs 8, 11 and 22 have an average potential traversal time of 7×10^{-10} sec, and the average of their ratios x'/x_0' is 0·54. Thus there is no possibility of making a lifetime determination by considering the few important events alone. In future work it is clearly necessary to improve the selection of decays with long values of T.

The average value of x'/x_0' is influenced by the correction procedure used when measuring x' and x_0'. That the average of x'/x_0' depends on the choice of the correction length is shown in table 3, where the correction procedure is indicated in column (1), and the corresponding average ratio is given in column (2). The amount of data diminishes as the correction lengths are increased, as shown in column (3).

Table 3. The Influence of the Correction Procedure

(1) Correction procedure	(2) Average ratio	(3) No. of decays
Full x, x_0	0·46	19
x —3 cm, x_0 —4 cm	0·38	18
x —5 cm, x_0 —6 cm	0·38	14

The average ratios in column (2) become approximately constant after 3 cm have been subtracted from each value of x and 4 cm from each x_0, and thus the bias effects are probably eliminated by these reductions.

A statistical treatment of lifetime data in the form of values of t, within a potental traversal time T, was given by Peierls (1936). He showed that a minimum value of the lifetime, independent of T, is given by the arithmetic mean of the values of t. The average of t, calculated from the data given in column (9) of table 1, is $1{\cdot}0\times10^{-10}$ sec. The statistical theory given by Peierls was discussed by Hole (1947) and a maximum likelihood procedure for finding the most probable lifetime from a set of values of t and T was given by Bartlett (1936). The available data on $V^{\pm}$-particles is not sufficiently extensive to warrant a calculation of this type.

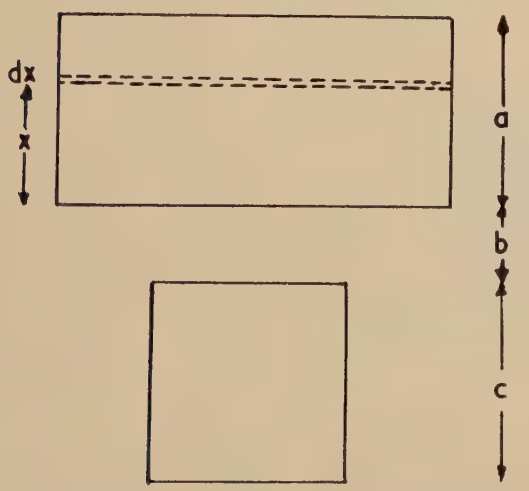
§ 4. The Frequency of Production of Charged V-Particles

Assuming that the $V^{\pm}$-particles observed in the cloud chamber are predominantly of the same type it is possible to find an approximate relationship between the frequency of production and various assumed lifetimes. An idealized version of the Pic-du-Midi apparatus, shown in fig. 1, is used and the following simplifying assumptions are made. The production of $V^{\pm}$-particles is assumed to occur uniformly throughout the lead absorbers, and the fraction of the shower particles which consists of $V^{\pm}$-particles is regarded as independent of the energy of the primary particle. Attention is restricted to shower particles with momenta greater than 10^9 ev/c and with inclinations to the vertical of less than 30°. The inclinations of the shower particles in the producing layer are neglected, and it is assumed that they have a path length in the cloud chamber equal to the average value of 15 cm. Furthermore, in this high momentum range, it can be assumed that the momentum spectrum follows an inverse power law. A negative exponent of three is used for the differential momentum spectrum of all the shower particles (see Barker and Butler 1951).

The ratio of the number, N_V, of $V^{\pm}$-particles produced to the total number of shower particles, N_s, is $F=N_V/N_s$, which can be written $(N/N_s)/(N/N_V)$, where N is the number of V-particles decaying in the chamber. There were 8000 ($\pm 25\%$) tracks of shower particles and 14 V-particle decays, satisfying the requirements laid down in the initial assumptions. Thus N/N_s is equal to $1{\cdot}75\times 10^{-3}$. The ratio N/N_V can be calculated as a function of τ_0, the assumed lifetime of the V-particles viz.:

$$\frac{N}{N_V}=\frac{\int_{P_0}^{\infty}\int_0^a [\exp-\alpha(x+b)][1-\exp(-\alpha c)]P^{-3}\,dP\,dx}{\int_{P_0}^{\infty}\int_0^a P^{-3}\,dP\,dx}, \quad . \quad (1)$$

Fig. 1



Simplified arrangement of cloud chamber and lead absorber

where $\alpha=M/P\tau_0$, M and P being the mass and momentum of a V-particle. The integrations are performed over the thickness of the lead absorbers, a, and for particles with momenta greater than P_0. The relationship between F and τ_0 is shown in fig. 2 for the following conditions: $P_0=10^9$ ev/c, $a=15$ cm, $c=15$ cm, $b=5$ or 10 cm, and $M=1\,200\,m_e$.

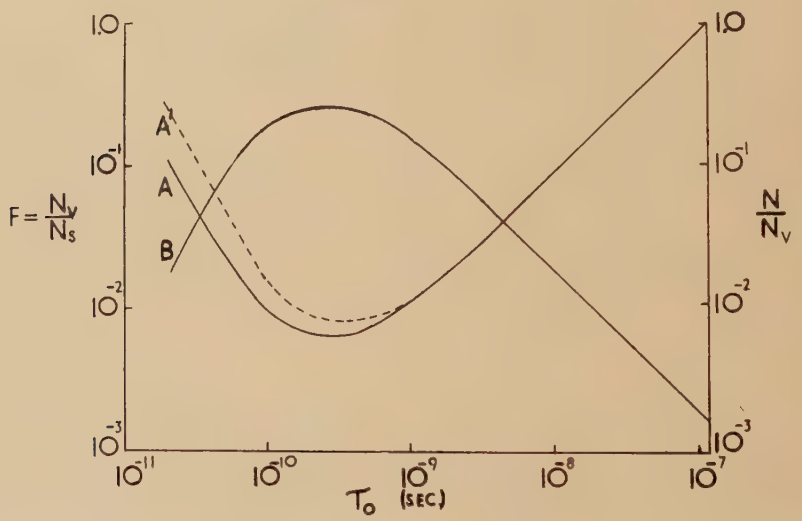
The various simplifying assumptions that have been made in these calculations are not very critical; for example, increasing the gap between the absorbers and the chamber from 5 cm to 10 cm does not make a very large difference to the frequency-lifetime curve.

The estimated average energy of the primary particles incident on the absorbing layer which trigger the counter detection system is 20–40 Bev. The average energy, however, of all interactions contributing to the flux of shower particles with momenta greater than 10^9 ev/c is probably

somewhat lower than this value since a considerable number of secondary interactions occur in the absorber. Thus the curve in fig. 2 refers to $V^{\pm}$-particles produced in showers of average energy about (10–30) Bev. If the V-particles were to interact in the lead above the chamber with the geometric cross section the values of N_V/N_s in fig. 2 would be increased by about a factor of two.

Data obtained in photographic emulsion work (Perkins 1952) make it likely that particles of mass about 1 200 m_e constitute less than 10% of the total number of shower particles in penetrating showers of average energy about 15 Bev. Using this value, an upper limit of about 10^{-8} sec is found for the mean lifetime of the charged V-particles which are observed to decay in the cloud chamber. The minimum frequency of production of about 1% corresponds to a mean lifetime in the range 10^{-10}–10^{-9} sec.

Fig. 2



The relations between the lifetime and frequency of production of $V^{\pm}$-particles.

(i) Curves A and A′ show N_V/N_s as a function of τ_0 for distances between the absorber and the chamber of 5 and 10 cm respectively.

(ii) Curve B shows N/N_V as a function of τ_0 for a distance of 5 cm between the absorber and the chamber.

§ 5. Conclusions

Under the assumption that the charged V-particles decaying in flight in the cloud chamber are predominantly of one type and similar to the κ-particles observed in photographic emulsions (Armenteros *et al.* 1952), limits can be placed on the mean lifetime of the particles.

Following the suggestions of Wilson and Butler (1952) data concerning the distribution across the chamber of the decays in flight of eighteen $V^{\pm}$-particles, observed in a chamber 28 cm in diameter, have been

presented. A minimum value of 10^{-10} sec has been found for the mean life. Considerably more data are required before a precise estimate can be made.

The frequency of production of $V^{\pm}$-particles has been considered as a function of various assumed lifetimes. From the recent photographic plate experiments, which show that κ-particles of mass about $1\,200\,m_e$ constitute less than 10% of the total number of shower particles, an upper limit of 10^{-8} sec is found for the mean lifetime.

These results may be compared with those obtained by other workers. Daniel *et al.* (1952) suggest that the lifetime of κ-particles is about 10^{-10} sec. However, estimates obtained from cloud-chamber investigations of slow unstable particles yield a lower limit of 10^{-9} sec (Astbury *et al.* 1952, Annis *et al.* 1952). The results obtained in the present work, concerning the energetic $V^{\pm}$-particles produced in interactions of average energy 10–30 Bev, are consistent with either of these values.

Acknowledgments

The authors are very much indebted to Professor P. M. S. Blackett for the benefit of his advice. We have also had helpful discussions with Professors M. S. Bartlett and J. G. Wilson. Thanks are also due to Dr. J. Rösch, the Director of the Observatoire du Pic-du-Midi, who has continued to give us every possible assistance at his laboratory. A financial grant has been received from the Department of Scientific and Industrial Research, which has also awarded a grant to M. G. Sowerby. C. M. York has received a grant from the United States Education Commission in the United Kingdom.

References

Annis, M., Bridge, H. S., Courant, S., Olbert, S., and Rossi, B., 1952, private communication.
Armenteros, R., Barker, K. H., Butler, C. C., Cachon, A., and York, C. M., 1952, *Phil. Mag.*, **43,** 597.
Astbury, J. P., Chippindale, P., Millar, D. D., Newth, J. A., Page, D. I., Rytz, A., and Sahiar, A. B., 1952, private communication.
Barker, K. H., and Butler, C. C., 1951, *Proc. Phys. Soc.* A, **64,** 4.
Bartlett, M. S., 1936, *Proc. Roy. Soc.* A, **154,** 124.
Bridge, H. S., and Annis, M., 1951, *Phys. Rev.*, **82,** 445.
Daniel, R. R., Davies, J. H., Mulvey, J. H., and Perkins, D. H., 1952, *Phil. Mag.*, **43,** 753.
Hole, N., 1947, *Arkiv für Matematik, Astronomi och Fysik*, Band 3B, No. 12.
O'Ceallaigh, C., 1951, *Phil Mag.*, **42,** 1032.
Peierls, R., 1935, *Proc. Roy. Soc.* A, **149,** 467.
Perkins, D. H., 1952, Report at Copenhagen Conference in June.
Powell, C. F., 1952, Report at Copenhagen Conference in June.
Wilson, J. G., and Butler, C. C., 1952, *Phil Mag.*, **43,** 993.

CXX. *Screw Dislocations, Etch Figures, and Holes*

By F. Hubbard Horn
General Electric Research Laboratory, Schenectady, New York*

[Received June 10, revised July 30, 1952]

Abstract

Etch figures similar in appearance to those observed in the mineralogical technique are shown to result from the rapid chemical dissolution of growth spirals on silicon carbide. Each spiral resulting from an independent screw dislocation gives an etch figure. The etch figure produced by rapid dissolution, therefore, becomes a convenient means for indexing screw dislocations whether growth spirals are evidenced or not. Prolonged rapid dissolution of a crystal with a screw dislocation extending between opposite crystal faces results in a hole through the crystal at the site of the screw dislocation.

Small round holes in crystals of AlB_2 formed in an aluminium melt and recovered by dissolving away the melt with strong acid have been described (Horn, Kasper and Fullam 1952). Spiral growth was observed on the AlB_2 crystals prior to the formation of the hole. Treatment of the crystals in strong acid was found a necessary part of producing the hole. The site of the hole is the centre of the spiral. Since the time F. C. Frank originally proposed that crystal growth may be initiated by screw dislocations (Frank 1949), A. R. Verma (1951) among others has successfully demonstrated that spiral growth patterns are evidence for growth by a screw dislocation mechanism. The holes in AlB_2 crystals have, therefore, been attributed to the preferential chemical attack at a screw dislocation.

Silicon carbide has been investigated more thoroughly for spiral growth than other substances. It seemed desirable to extend the study of the rapid dissolution of growth spirals to include those on SiC. Consideration of the process of dissolution of growth spirals leads to the conclusion that in order to attack the region of the screw dislocation preferentially, the process must be rapid. Dissolution near equilibrium conditions will cause a general disintegration of the crystal. The difference is readily demonstrated in the case of the AlB_2 crystals depending on the concentration of acid used. The emphasis of these studies is on the results from rapid dissolution.

Silicon carbide is attacked by alkali carbonate fusion (Scott 1939). If the reaction is carried out in air at about 1000° C, there appears to be a uniform attack of silicon and carbon atoms. At lower temperatures,

---

* Communicated by the Author.

a carbon residue may adhere to the crystals. The details of the results obtained depend somewhat on the carbonates used. However, the conclusions drawn from the present studies are sufficiently qualitative that they are not influenced by the detail of the etching process beyond knowing that the etching has been rapid.

After selecting silicon carbide crystals with the desired spiral growth patterns, the procedure for following the dissolution of the growth patterns consisted of making photographs of replicas prepared from the crystal surface after each stage of alkali carbonate fusion.

In Plate LXXIII there are shown photographs of replicas of the surface of a SiC crystal selected because the entire surface (*a*) was marked by terraces resulting from the interaction of eight spirals, (*b*) of the same sense. Using the criterion that the step width per revolution for each spiral is constant, each spiral may be considered to be derived from an independent dislocation. In the photographs (*c*) through (*f*) there are shown replicas of the same area as in (*b*) after successive treatments of the crystal in a fused alkali carbonate mixture at approximately 1000° C for the total times indicated. After two minutes, the spirals are obliterated. The centre of each spiral becomes indexed by a very fine pit, which on successive treatments enlarges and deepens. The underlying hexagonal symmetry is eventually apparent from the rays within the etched figures.

Although one must keep in mind that the detailed structure of etched figures depends on the etching conditions, it is seen from the above results that the centre of each isolated spiral arising from an independent screw dislocation is through rapid etching replaced by a pit or etch figure. Thus, the etch figure may, under such conditions, be used to identify the centre of a spiral or more particularly, to identify the position of a screw dislocation.

From the results obtained with independent dislocation spirals, one may turn to a crystal known to be made up from many spirals, or multiple interacting dislocations. Plate LXXIV (*a*) shows a photograph of a replica of the surface of a SiC crystal seen to exhibit several areas and kinds of growth spirals. In (*c*) is shown enlarged the spiral to the right in (*a*). It appears to begin from a single dislocation and after several turns, merges with many small spirals to one side. In (*b*) is shown enlarged the area to the left in (*a*). This area contains two spirals composed of the interaction of a number of apparently closely spaced dislocations. Photographs in Plate LXXV show the crystal area in Plate LXXIV (*a*) after treatment in fused Na_2CO_3 for the times indicated.

Several features may be noted from these results. As previously, the centre of the independent dislocation spiral is replaced by an etch pit. The centre of each complex spiral is also replaced by an etch pit. In addition, however, there are many new etch pits—etch pits for which there was not a visible corresponding surface spiral. Another striking feature of the etch pits is their arrangement. These are, in general, arranged in rows between which are the 30-, 60-, 120- and 90-degree angles

characteristic of the underlying hexagonal symmetry of the crystal in the plane of the face. One may only speculate concerning the significance of the arrangement of the etch figures. These linear arrays may be evidence of dislocations caused by deformation of the crystal during growth. Such deformation has been proposed by F. C. Frank (1951) and is used by A. J. Forty (1952) to account for the large step heights encountered in spiral growth on CdI_2. Alternatively, the rows of figures may be connected with adsorbed impurity or dislocations formed at the edges of the crystal during growth. Another feature of the etch figures, seen particularly in Plate LXXV (*a*), (*b*), and (*c*) is that they are arranged approximately evenly spaced in the rows, rather as though a repulsive force existed between them.

One may now turn to see what happens upon etching a crystal with no visible growth spirals on the surface, but which is obtained from the same group of crystals among which there are those showing spiral growth. The photographs in Plate LXXVI are of the surface of SiC crystals etched as indicated. Although no spirals were detected on the surfaces of these crystals, etching produces a large number of etch figures similar in appearance to those observed from crystals with spirals. One derives the impression from looking at many of these crystals that there are, in fact, many more etch figures obtained from crystals with no observable spirals than from crystals on which spiral growth is observed. This is consistent with the idea that such crystals may have smooth surfaces because of the interaction of the growth steps from a large number of screw dislocations. Similarly to the previous two situations, the etch figures are arranged in rows which display the hexagonal symmetry in the face plane and being approximately evenly spaced in the rows suggest a repulsive force between dislocations.

It is of interest now to see what happens upon prolonged etching. Though more complex situations are geometrically possible, it is expected that frequently a screw dislocation extends perpendicularly through from face to face of a thin crystal. Then, upon prolonged etching, a hole may be produced through a crystal. This process was proposed for the holes in AlB_2. That the same is true for SiC is seen in the photographs in Plate LXXVII which show holes produced in SiC by alkali-carbonate fusion. For preparing photograph (*c*), gold was evaporated on the surface of the crystal in order to improve the contrast. When small, the holes are round ; when etched to a large size, the hexagonal symmetry becomes apparent.

The results reported here appear to have much in common with the technique of producing etch figures for interpreting the symmetry of mineralogical specimens (Honess 1927), thus prompting some speculation. In common is the fact that the results depend on the etchant and the rate of etching, that in both cases the crystal symmetry is revealed by the figure developed, and that the figures themselves generally are arranged according to the crystal symmetry. The deeply etched rows of screw

dislocation sites have the appearance of 'solution channels' often encountered in the mineralogical technique. Thus, one is led to think that the origin of many etch figures on natural minerals may be due to the presence of screw dislocations in the mineral. A corollary to this point of view is that natural minerals should be a source of materials showing spiral growth. It was on natural beryl that spiral growth was first recognized (Griffin 1951). Spiral growth has also been found on natural crystals of barite (Horn, unpublished), graphite (Horn 1952), and micas (Amelinckx 1952). A reinvestigation of minerals from the viewpoint of screw dislocations may be rewarding in understanding the properties of screw dislocations and, also, in providing additional evidence for the growth conditions in various deposits.

Although the effect of the results reported here is to raise rather than to answer problems, the study is reported now because there is implied in the results a technique for revealing screw dislocations in crystals. The technique depends on the differential rates of chemical attack of a screw dislocation and good crystal when the etching is rapid. The technique appears to be applicable whether there is evidence of screw dislocations from spiral growth patterns or not and may, therefore, be presumed to apply to revealing screw dislocations emerging from a surface whether the dislocations are introduced subsequent to or during crystal growth.

The author is pleased to acknowledge the benefits from discussion with and the interest of Dr. J. S. Kasper, Dr. M. H. Hebb, and Dr. L. Apker of this laboratory. The criticisms of Dr. F. C. Frank during the preparation of the manuscript are also gratefully acknowledged.

References

Amelinckx, S., 1952, *Nature, Lond.*, in press (cf. 1952, *Comptes Rend.*, **234,** 113).
Forty, A. J., 1952, *Phil. Mag.*, **43,** 72.
Frank, F. C., 1949, *Trans. Faraday Soc., Disc. No.* 5; 1951, *Phil. Mag.*, **42,** 1014.
Griffin, L. J., 1951, *Phil. Mag.*, **42,** 775.
Honess, A. P., 1927, *Etch Figures on Crystals* (New York: J. Wiley and Sons).
Horn, F. H., 1952, *Nature, Lond.* (in press).
Horn, F. H., Kasper, J. S., and Fullam, E. F., 1952, *Nature, Lond.* (in press).
Scott, W. W., 1939, *Standard Methods of Chemical Analysis*, 5th Ed. (New York: D. van Nostrand and Co.).
Verma, A. R., 1951, *Phil. Mag.*, **42,** 1005.

CXXI. CORRESPONDENCE

Angular Correlations in the Reaction $^{15}N(p, \alpha\gamma)^{12}C$

By J. Seed and A. P. French
Cavendish Laboratory, Cambridge*

[Received August 25, 1952]

The bombardment of ^{15}N by protons has been studied by Schardt *et al.* (1952), who found several resonant states of ^{16}O for which the chief mode of decay is to the 4·47 Mev level of ^{12}C by emission of low energy alpha-particles.

We have studied the angular correlation between these alpha-particles and the subsequent gamma-rays at the 898 kev resonance, which is sufficiently strong and isolated to be regarded as a pure state.

Targets of ^{15}N were obtained from the Atomic Energy Research Establishment, Harwell; they were formed by bombarding thin aluminium sheets in the electromagnetic separator. An aluminium backing was chosen in order to reduce the number of elastically scattered protons from the target.

In our experiments, alpha-particles emitted at 90° to the incident proton beam were separated from the scattered protons by magnetic analysis, and were detected by a thin ZnS screen and E.M.I. photomultiplier. The gamma-rays emitted in the reaction were detected by a $1\frac{1}{2}$ in. $\times 1\frac{1}{2}$ in. mosaic of NaI crystals and a photomultiplier. The crystals were situated some 2 in. from the target, giving a solid angle of about $\frac{1}{30}$ sphere, and their detection efficiency was about 30% for the 4·47 Mev gamma-rays. By rotation of the gamma-phosphor about the target, the (α, γ) correlation could be studied over the angular range 90° to 180°. The arrangement was such that an (α, γ) correlation could be studied in a plane either containing or perpendicular to the incident proton direction.

It is intended to give a fuller description of the apparatus elsewhere.

Both the alpha and gamma single counting rates were continuously monitored, and the number of coincidences was obtained as a fraction of the number of alpha-particles counted at a given angle. The coincidence resolving time was 0·3 μsec.

The experimental points are shown in fig. 1, and it is seen that a high order of anisotropy is present in the (α, γ) correlations in both planes. The results can be fitted if one assumes that the incoming protons have orbital momentum 2 and that the compound nucleus $^{16}O^*$ is (2—) (i.e. has spin 2 and odd parity). There is evidence to show that the first

* Communicated by the Authors.

excited state of ^{12}C is (2+) (Lewis 1952, Hubbard *et al.* 1952, Kraus and French 1952); it could thus be formed from the $^{16}O^*$ by emission of alpha-particles with $l=1$.

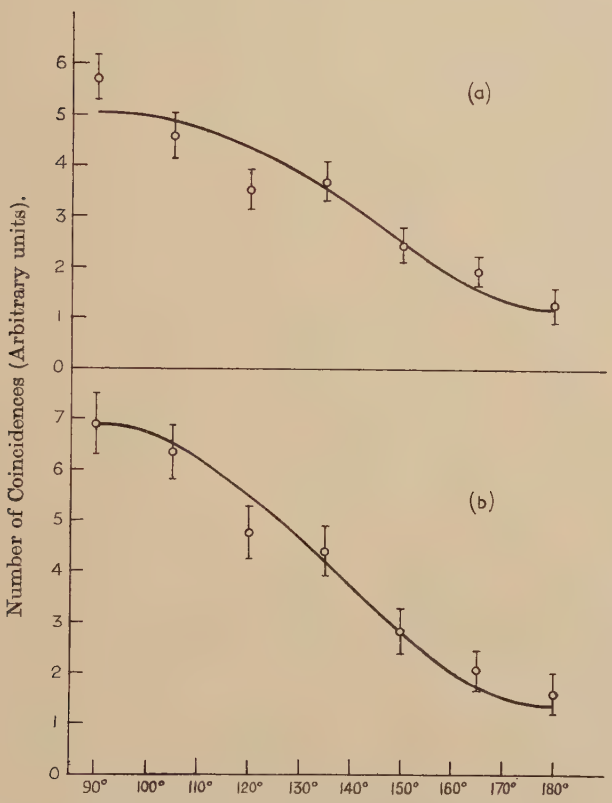
On the above assumption, the theoretical correlation functions are

$$I\left(\theta, \frac{\pi}{2}\right) = 1 - \frac{6}{6+f}\cos^2\theta$$

for the (α, γ) correlation in a plane perpendicular to the incident proton direction and

$$I(\theta, 0) = 1 + \frac{5f-6}{6}\cos^2\theta - \frac{2f}{3}\cos^4\theta$$

Fig. 1



(α, γ) angular correlations in $^{15}N(p, \alpha\gamma)$ at 898 kev: (*a*) in plane containing proton beam; (*b*) in plane perpendicular to beam. The alpha particles were observed at 90° to the beam in each case. Errors shown are standard deviations.

for the correlation in a plane containing the incident proton direction. In these expressions, f is the relative probability of forming the ^{16}O compound nucleus from the parallel and anti-parallel spin configurations respectively of the colliding particles.

The full curves of fig. 1 are obtained by putting $f=0{\cdot}72$ and correcting for the finite geometry of the apparatus. The value of f is deduced from the observed angular distribution of the gamma-rays relative to the proton beam at 898 kev (Kraus and French 1952), and it can be seen that an adequate fit to the experimental points is obtained. One should note that alpha-particles with $l=3$, as well as $l=1$, might be emitted from a (2—) state of $^{16}O^*$; it was not, however, found necessary to consider these in fitting either the (p, γ) distribution or the (α, γ) correlations.

We are grateful to the staff of the Atomic Energy Research Establishment, Harwell, for providing separated nitrogen isotope targets. One of us (J. S.) is indebted to the Department of Scientific and Industrial Research for a grant.

References

Hubbard, T. P., Nelson, E. B., and Jacobs, J. A., 1952, *Phys. Rev.*, **87**, 378.
Kraus, A. A., and French, A. P., 1952, to be published.
Lewis, G. M., 1952, *Phil. Mag.*, **43**, 690.
Schardt, A., Fowler, W. A., and Lauritsen, C. C., 1952, *Phys. Rev.*, **86**, 527.

An Effect of Electron Bombardment on Order in Cu_3Au *Alloy*

By J. Adam, A. Green and R. A. Dugdale
Atomic Energy Research Establishment, Harwell, Berks.*

[Received September 8, 1952]

The experiment described in this letter was designed in order to study the part played by lattice defects, introduced by high energy electron bombardment, in the ordering of a Cu_3Au alloy.

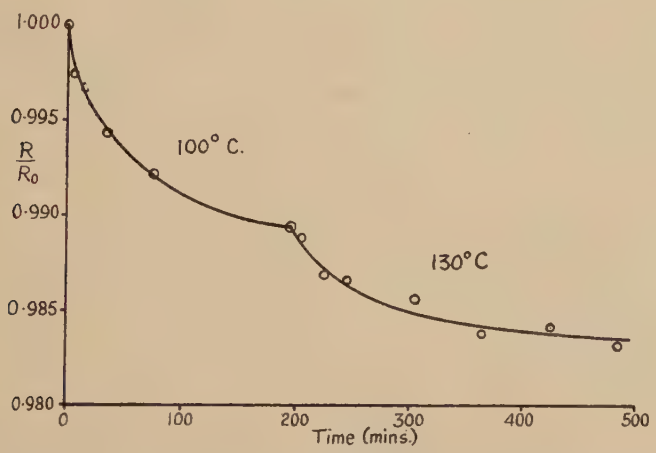
The effect of a pile irradiation on the electrical resistance and unit-cell dimensions of an ordered Cu_3Au alloy containing 25·8 atomic per cent Au was studied recently by Adam and Dugdale (1951). Their main conclusion was that, under suitable conditions, neutron bombardment is capable of producing simultaneous ordering and disordering effects. Using an initially disordered Cu_3Au resistor, Blewitt and Coltman (1952) found that the relaxation rate for ordering at 200° c was considerably increased by neutron bombardment in a pile. They attributed this effect to the accelerated growth of ordered domains by the diffusion of Frenkel effects formed by the bombardment.

In the present experiment a resistor of the 25·8% alloy was prepared in the form of a strip 0·0015 in. thick. After annealing to remove the effects of cold work a high degree of order was induced in it by heating

---

* Communicated by the Authors.

for six days at 381° C (which is about 7° C below the measured critical temperature), followed by slow cooling to room temperature at 20° C per hour. After this treatment the resistivity at 0° C was about 4·5 ohms cm^{-1}. X-ray diffraction photographs of similarly treated specimens show strong and sharp superlattice lines which indicate a high degree of order. The specimen was then bombarded for 2 hours with 1 MeV electrons, the integrated flux being about 3×10^{17} electrons/sq. cm, the specimen temperature varying between 40° C and 45° C. The bombardment caused no measurable change in resistance. The specimen was then heated at 100° C for various periods of time; its resistance (measured at 0° C) decreased as shown in the graph. Subsequent heating at 130° C resulted in a further decrease in resistance. Further electron bombardment of the specimen followed by heating at 100° C, again produced a marked decrease in resistance, and a third irradiation resulted in a similar effect. Preliminary experiments with 3 MeV electron bombardments gave similar results.



Fractional change in resistance v. annealing time, at 100° C and 130° C, for an ordered Cu_3Au resistor after electron bombardment (1 MeV).

On the other hand, heating the specimen for 3 hours at 100° C before the first bombardment produced no measurable change in resistance.

1 MeV electrons are capable of giving up to 68 ev recoil energy to a Cu atom, and up to 22 ev to a Au atom in elastic collisions with the nuclei. Seitz (1949) has suggested a nominal figure of 25 ev for the energy required to form an interstitial atom by bombardment. Assuming this to be the case in Cu_3Au, the electrons are capable of knocking at least the Cu atoms out of their normal sites into interstitial positions, forming the so-called Frenkel defects. The cross section for 25 ev or more to be given to a Cu atom is of the order of 20 barns (estimated from the formula given by Mott and Massey 1949). Thus, the electron bombardment may be expected to have created Cu Frenkel defects with

a concentration of about 6 per million. If one or other of the components of these defects—viz. interstitial atoms or atomic vacancies—acquired sufficient mobility at 100° C to diffuse some distance through the lattice before being captured, movement of some Cu and Au atoms from 'wrong' to 'right' positions could take place. The increase in order thus produced would explain the observed decrease in resistance.

According to the theory of Sykes and Jones (1936) the specimen here used would initially contain large domains within which the equilibrium degree of order would correspond to a temperature in the neighbourhood of 250° C. When large domains are present the electrical resistance is not very dependent on domain size and it would seem most probable therefore that the diffusing lattice defects increased the order within the domains. The experiment described here suggests that by introduction of artificially produced lattice defects it is possible to reach an equilibrium degree of order corresponding to a temperature lower than that obtainable by purely thermal treatment in a reasonable time.

The authors' thanks are due to Messrs. T. M. Fry and J. H. W. Simmons for stimulating discussions on this subject, to Messrs. E. R. Wiblin and J. D. Milne and members of the Linear Accelerator and Van der Graaf sections for electron irradiations, and to the Director, Atomic Energy Research Establishment, for permission to publish this letter.

References

Adam, J., and Dugdale, R. A., 1951, Reported in *Nature, Lond.*, **168,** 581.
Blewitt, T. H., and Coltman, R. R., 1952, *Phys. Rev.*, **85,** 384.
Mott, N. F., and Massey, H. F. W., 1949, *The Theory of Atomic Collisions*, 2nd Editions (Oxford : University Press), p. 80.
Seitz, F., 1949, *Disc. Far. Soc.*, p. 271.
Sykes, C., and Jones, F. W., 1936, *Proc. Roy. Soc.*, **157,** 213.

Critical Shear Stress and Temperature

By E. N. da C. Andrade, F.R.S.*

[Received September 8, 1952]

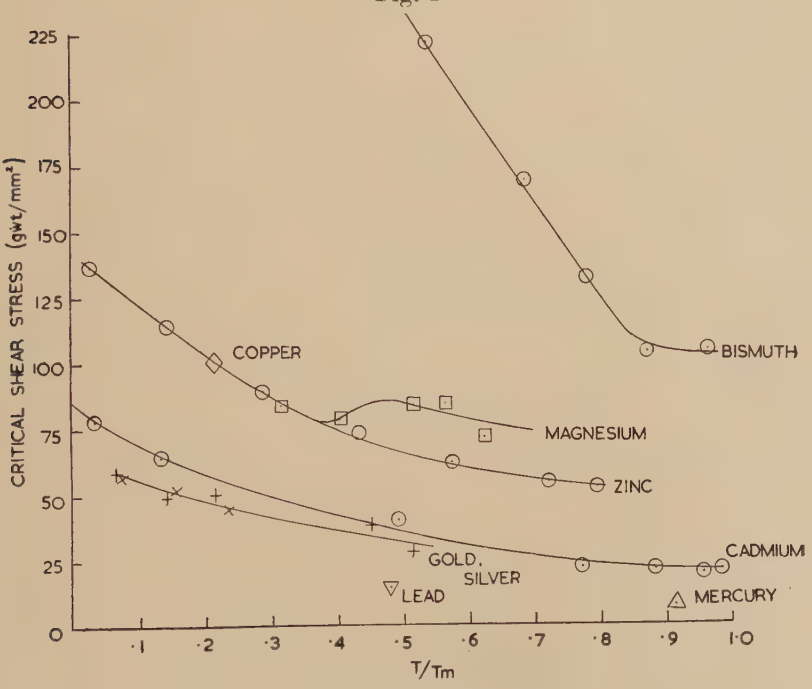
It is generally accepted that, as regards the mechanical behaviour of single crystals, there is a contrast between the metals of hexagonal structure which, in general, possess one set of glide planes, and those, like the cubic metals, with a multiplicity of glide planes. This view is supported by consideration of the curves showing the variation of resolved shear stress with glide over a wide range of glide. The diagram originally prepared by E. Schmid (1934) clearly separates the two classes, the cubic metals showing very much greater hardening with glide. It is

* Communicated by the Author.

to be noted that the group of cubic metals (Al, Ni, Cu, Ag, Au) is also the group of relatively high melting point, but even when allowance is made for this the distinction still exists.

However, Dr. Henderson and I (1951) have recently found that the pure f.c. cubic metals, gold and silver, show the hexagonal type of glide ('easy glide') at low temperature when the glide is not too large. Röhm and Kochendorfer (1950) have shown that a single crystal of aluminium (f.c. cubic), subjected to approximately simple shear, gives the hexagonal type of stress–strain curve as against the type normally found with cubic metals in tension and, in particular, that found by other workers with aluminium under the ordinary test conditions. The difference between the mechanical behaviour of the two classes of metals appears, then, to be less fundamental than has been suggested. With suitably chosen conditions glide in cubic metals can apparently be restricted to one set of glide planes, without even local interference by other sets, and then the behaviour is of a hexagonal nature.

Fig. 1

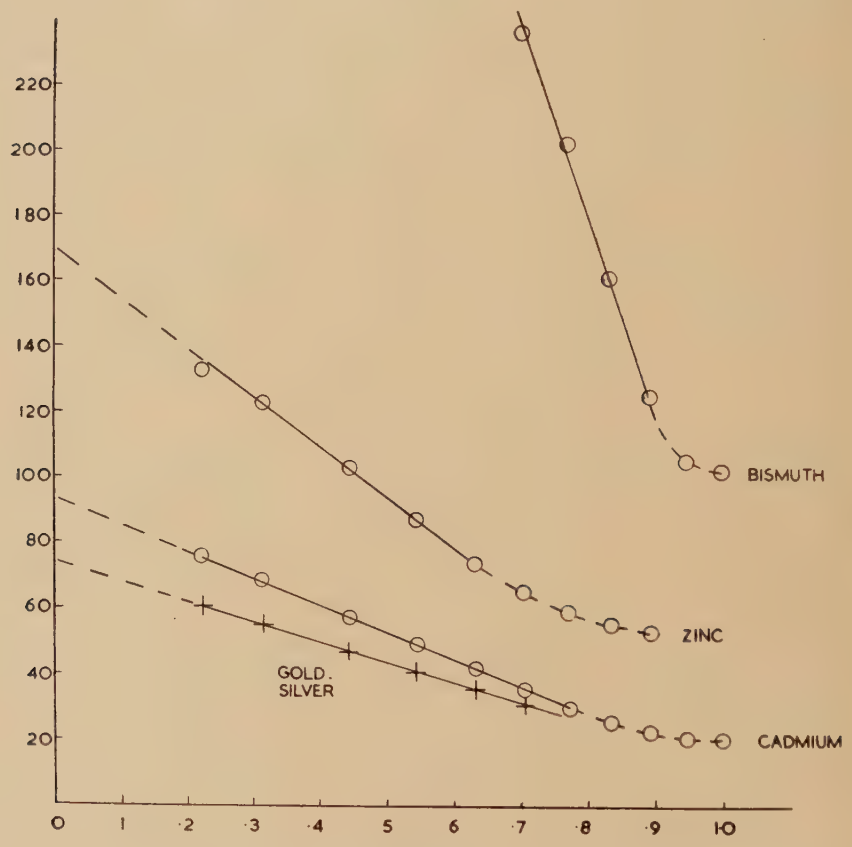


Critical shear stress against $\theta = T/T_m$

⊙ : bismuth, zinc, cadmium
◇ : copper
□ : magnesium
× : silver
\+ : gold
▽ : lead
△ : mercury

The critical shear stress shows no distinction between the two types of metals. Schmid, in the paper quoted, gives a diagram of critical shear stress against temperature for four metals (Cd, Zn, Mg, Bi), but no systematic comparison with cubic metals appears to have been made. It has been suggested (Andrade 1938, Andrade and Chow 1940) that in comparing the mechanical properties of different metals it is advisable to take as the measure of the temperature the ratio $\theta = T/T_m$, where T and T_m are respectively the temperature in question and the temperature

Fig. 2



Critical shear stress against $\theta^{1/2}$. The points are taken from the smoothed curves in fig. 1.

of melting, both absolute. Accordingly I have prepared a diagram (fig. 1) showing the variation of critical shear stress with θ for gold and silver (Andrade and Henderson), cadmium (Boas and Schmid), magnesium (Schmid 1931), zinc (Fahrenhorst and Schmid) and bismuth (Georgieff and Schmid) and have inserted single values for lead (Makin, unpublished), copper (Sachs and Weerts) and mercury (Greenland). The critical shear

stress is notoriously a quantity which is very structure-sensitive, but the values represented in the diagram are for pure metals and seem consistent and reliable. For aluminium the values are so variable that I have not included the metal.

It will be seen that the curves for the different metals show a very similar course, which they do not do if the critical shear stress σ is plotted against T. The values for the different crystal types show no suggestion of grouping according to crystal structure. With magnesium, which is hexagonal and normally glides on the basal planes, pyramidal glide on $(10\bar{1}1)$ faces is established at about 225° C ($\theta=0{\cdot}54$), which accounts for the peculiar form of the curve for this metal.

According to Orowan (1934) the variation of critical shear stress with T is given by

$$\sigma=\sigma_0-BT^{1/2}.$$

In fig. 2, σ is plotted against $\theta^{1/2}$, the points being taken from the smooth curves in fig. 1 (which were drawn before this plot was made). For gold, silver, cadmium and zinc Orowan's law holds well within the range $\theta=0{\cdot}05$ to $0{\cdot}4$, although for higher temperature the experimental values are higher than it would indicate. For bismuth values are not available for low θ's.

References

Andrade, E. N. da C., 1938, *Proc. Roy. Soc.* A, **168,** 310.
Andrade, E. N. da C., and Chow, Y. S., 1940, *Proc. Roy. Soc.* A, **175,** 290.
Andrade, E. N. da C., and Henderson, C., 1951, *Phil. Trans. Roy. Soc.* A, **244,** 177.
Boas, W., and Schmid, E., 1929, *Z. f. Physik*, **57,** 575.
Fahrenhorst, W., and Schmid, E., 1930, *Z. f. Physik*, **64,** 845.
Georgieff, M., and Schmid, E., 1926, *Z. f. Physik*, **36,** 759.
Greenland, K. M., 1937, *Proc. Roy. Soc.* A, **163,** 34.
Orowan, E., 1934, *Z. f. Physik*, **89,** 605.
Röhm, F., and Kochendorfer, A., 1950, *Z. f. Metallkunde*, **41,** 265.
Sachs, G., and Weerts, J., 1930, *Z. f. Physik*, **62,** 473.
Schmid, E., 1931, *Z. f. Elektrochem*, **37,** 447.
Schmid, E., *International Conference on Physics*, 1934 (Physical Society, 1935), Vol. II, p. 161.

The Sintering of Powders and Diffusion

By R. S. Barnes
Atomic Energy Research Establishment, Harwell*

[Received August 15, 1952]

It is the purpose of this communication to apply the result of some recent interdiffusion experiments (Barnes 1952), using solid sandwiches of metals which form a complete series of solid solutions, to the density

---

* Communicated by the Author.

behaviour of mixed powders which also form solid solutions, and to the more fundamental problem of the mechanism of the sintering process.

It has been observed that the density of mixed copper and nickel powders decreases initially (by sometimes as much as 20%) before it starts to increase on heating at temperatures where interdiffusion can occur (Rhines and Meussner 1943, Alexander 1951, Butler and Hoar 1952). Changes in lattice parameter cannot account for this large decrease, nor can the expansion of gas entrapped during compacting, as compacts consisting of copper powder only do not behave similarly (Butler and Hoar 1952).

Density changes similar to these have been observed when some mixed spinel powders are heated together (Rigby 1952).

Diffusion experiments have shown that voids form on the copper-rich side of the marked join in copper–nickel couples (Seith and Kottmann 1952, Barnes 1952). Similar voids resulting from the interdiffusion of alpha–brass and copper, and aluminium–bronze and copper have been found by Bückle and Blin (1952). The formation of these voids in material which before diffusion had been quite solid reduced the density of the sandwiches by as much as 20% for almost complete diffusion (Barnes 1952). These experiments were able to show that the voids did not appear as a result of gas, either in the metals themselves or entrapped during the welding process. Thus in sintered *mixed* powders an extra volume increase, over and above any caused by included gas, must result from the formation of such voids, and it is suggested that this is the cause of the initial density decrease, and also the initial decrease in hardness and rapid rise in the electrical resistance observed by Rhines and Meussner (1943).

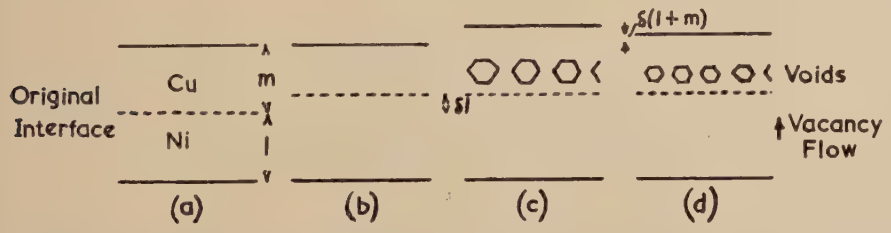
The voids resulting from the interdiffusion process have crystallographic shapes, therefore they should be distinguishable from voids formed by the closing of the pores between the original particles, which initially would have less regular shapes. A careful mechanical polish or microradiographic examination should show this.

The interdiffusion experiments can best be explained on the basis of a vacancy mechanism of diffusion where there is a resultant flow of vacancies from the nickel, where they are generated, to the copper, where they are lost. The vacancies can be lost in two ways, either by annihilation at lattice discontinuities, e.g. grain boundaries and edge dislocations, which results in a shift, relative to both copper and nickel surfaces, of the markers placed at the original interface ((*b*) in figure), or by coalescing to form voids, which results in a volume increase equal to the total volume of the voids and a shift of the markers relative to the nickel surface only (*c*). As the volume increases, and the interface moves relative to both the copper and nickel surfaces (*d*), both these processes must occur.

The number of vacancies coalescing to form voids is proportional to the volume increase ($\delta(l+m)$), and to $\sqrt{t}$ (where t is the annealing time) (Barnes 1952). The net number of vacancies which have crossed the

marked interface is proportional to δl (the shift of the markers relative to the nickel reference surface), and this is also proportional to $\sqrt{t}$. Thus the number of vacancies which have been annihilated at lattice discontinuities must also be proportional to $\sqrt{t}$. The relative shifts suggest that approximately equal numbers of vacancies must be lost by each of the two processes, and as they are both proportional to $\sqrt{t}$ it seems highly probable that a given proportion of the vacancies meet lattice discontinuities and are annihilated, the rest forming voids. Little sintering of the voids can occur during interdiffusion as this would be expected to give a dependence other than $\sqrt{t}$.

Thus, when interdiffusion causes the material to be *supersaturated* with vacancies, they are lost both by void formation and at lattice discontinuities, the energy needed to increase the area of the internal and external surfaces being derived from the process of mixing the two types of atom. However, when the material is of *homogeneous* composition (either because the original compact consisted of particles of only one



Representation of the cross-section of a couple (*a*) before the diffusion anneal, (*b*) after diffusion, all the vacancies haveng been annihilated, (*c*) all the vacancies going to form voids, and (*d*) what happens in practice.

material or because of complete diffusion) there will be a tendency to retain only the equilibrium number of vacancies, and these will not enlarge existing voids as there is no source of energy to increase the surface area. But vacancies will still be annihilated at lattice discontinuities without increasing the surface energy, locally reducing the number of vacancies in the lattice. This loss will be made up by a steady generation of vacancies at voids which will diminish in size as a result. Small voids would decrease in size more rapidly than large voids, because the gain in energy per vacancy generated is larger at small voids. Large voids might be expected even to grow at the expense of smaller voids because of the resultant decrease in surface energy, but the total volume of the voids would steadily decrease, and this was observed by Rhines, Birchenall and Hughes (1950). The steady annihilation of vacancies at lattice discontinuities will eliminate whole planes of atoms, causing a reduction in the total volume. Thus the sintering process probably consists of a special form of plastic deformation where the change in shape results from a flow of vacancies from the voids to lattice discontinuities.

References

Alexander, B. H., 1951, *J. of Metals*, **191,** 95.
Barnes, R. S., 1952, *Proc. Phys. Soc.* B, **65,** 512.
Buckle, H., and Blin, J., 1952, *J. Inst. Metals*, **7,** 385.
Butler, J. M., and Hoar, T. P., 1952, *J. Inst. Metals*, **5,** 207.
Rhines, F. N., Birchenall, C. E., and Hughes, L. A., 1950, *J. of Metals*, **188,** 378.
Rhines, F. N., and Meussner, R. H., 1943, *Symposium on Powder Metallurgy. Phil., American Soc. Testing Materials*, published in Philadelphia, U.S.A.
Rigby, G. R., 1952, private communication.
Seith, W., and Kottmann, A., 1952, *Die Naturwissenschaften*, **39,** 40.

The Temperature Variation of the Thermodynamic Potential of a Degenerate Electron Gas

By Sir K. S. Krishnan, F.R.S.
National Physical Laboratory of India, New Delhi
and
P. G. Klemens
Division of Physics, Commonwealth Scientific and Industrial Research Organization, Sydney*

[Received August 25, 1952]

In the Fermi–Dirac distribution function, namely

$$f(E)=\frac{1}{e^{(E-\zeta)/kT}+1}, \qquad \ldots\ldots\ldots \quad (1)$$

ζ, as is well known, is the thermodynamic potential, or the free energy at constant pressure, per electron, and is given by

$$\zeta=u-Ts+pv, \qquad \ldots\ldots\ldots \quad (2)$$

where the letters have their usual significance. Obviously ζ is a function of the temperature, and is determined by the equation

$$n=2\int_0^\infty N(E)f(E)\,dE, \qquad \ldots\ldots\ldots \quad (3)$$

where n is the number of electrons per unit volume, and $N(E)$ is the density of energy states. For an almost completely degenerate gas, i.e. at temperatures $T \ll T_0$, where

$$T_0=\zeta_0/k=\frac{h^2}{2mk}\left(\frac{3n}{8\pi}\right)^{2/3} \qquad \ldots\ldots\ldots \quad (4)$$

is the degeneracy-temperature, the integration of (3) yields to a first approximation

$$\zeta=\zeta_0(1-\gamma T^2/6), \qquad \ldots\ldots\ldots \quad (5)$$

where

$$\gamma=\pi^2k^2/2\zeta^2. \qquad \ldots\ldots\ldots \quad (6)$$

* Communicated by the Authors.

Since the second term in (5) is merely a correction term, ζ appearing in the denominator of (6) may be replaced by ζ_0, or by ζ_0^*, which we shall define presently, which are both of nearly the same magnitude as ζ.

In expression (5) ζ_0 is frequently referred to as the value of ζ at $T=0$, and for that reason is sometimes regarded as independent of T, and hence the temperature variation of ζ is taken to be determined completely by the second term, which involves T explicitly. It is the main purpose of this note to emphasize that this will be the case only if the density of electrons n is kept constant, whereas at constant pressure the density n, and hence also ζ_0 as will be seen from (4), will vary with the temperature. In other words ζ_0 will not represent the free energy at the same pressure at $T=0$. The latter energy will be given by

$$\zeta_0^*=\zeta_0(1+\gamma T^2/3), \quad \ldots\ldots\ldots \quad (7)$$

and the expression for the temperature variation of ζ at constant pressure will therefore be given by

$$\zeta=\zeta_0^*(1-\gamma T^2/2). \quad \ldots\ldots\ldots \quad (8)$$

Physically $\zeta_0^*-\zeta$ has the following significance. Let ϕ_T be the thermionic work function, as usually defined, of a metal at temperature T. If the electrons in the metal, i.e. in the condensed phase, can be regarded as an almost completely degenerate assemblage, having a finite latent heat of evaporation, then the temperature variation of ϕ *due to the temperature variation of* ζ will be given by

$$\phi_T-\phi_0=\zeta_0^*-\zeta. \quad \ldots\ldots\ldots \quad (9)$$

Incidentally it may be mentioned that thermodynamically

$$\zeta_0^*-\zeta=T\int_0^T \frac{c_p}{T}\,dT-\int_0^T c_p\,dT, \quad \ldots\ldots \quad (10)$$

where c_p is the specific heat of the electrons in the condensed phase at constant pressure. In view of (8) and (10) one obtains

$$c_p=\gamma T, \quad \ldots\ldots\ldots \quad (11)$$

which is the same expression as for c_v.

It should be mentioned here that if the electronic structure of the metal, instead of corresponding to a nearly empty parabolic band, as we have taken it to be till now, corresponds to a nearly full parabolic band, the expressions for c_p and c_v will remain the same as before, but will now refer to the holes, and hence $\zeta_0^*-\zeta$ in eqns. (8), (9) and (10) should be replaced by $\zeta-\zeta_0^*$. In other words ζ will now increase with the increase of the temperature, though the specific heats remain positive as they should.

That in a highly degenerate electron gas, the specific heat at constant pressure should be the same as that at constant volume can also be demonstrated otherwise. At higher temperatures naturally c_p will increase more rapidly than c_v, and they will tend to the values $\frac{5}{2}k$ and $\frac{3}{2}k$ respectively per electron.

CXXII. *Notices of New Books and Periodicals received*

Descartes and the Modern Mind. By Albert G. A. Balz. (New Haven: Yale University Press, 1952.) [Pp. 492.] Price $ 10.00.

This is an impressive essay in the history of ideas. The author shows how Descartes' philosophical position arises from the scholastic tradition as embodied in and modified by Thomas Aquinas. In developing his thesis Professor Balz generously acknowledges his indebtedness to Gilson and Koyré and gives good reasons for any disagreement with these scholars.

His mode of presentation will not be to everybody's taste. Thus, I believe that little help will be derived from the proposal to use " the name *Cartesius* to signify Reason itself " and to regard René Descartes " as one who proposes to report what he has learned from Cartesius, the very voice of Reason " (p. 67). It would be a pity if this and similar passages were to deter anybody from reading a very learned, well documented, and interesting contribution to the history of philosophy. S. K.

ERRATUM

On the Theory of Beta-Radioactivity, by A. M. Smith, 1952, *Phil. Mag.*, **43**, 915.

Figures 3 (*a*) and 3 (*b*) should be interchanged.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

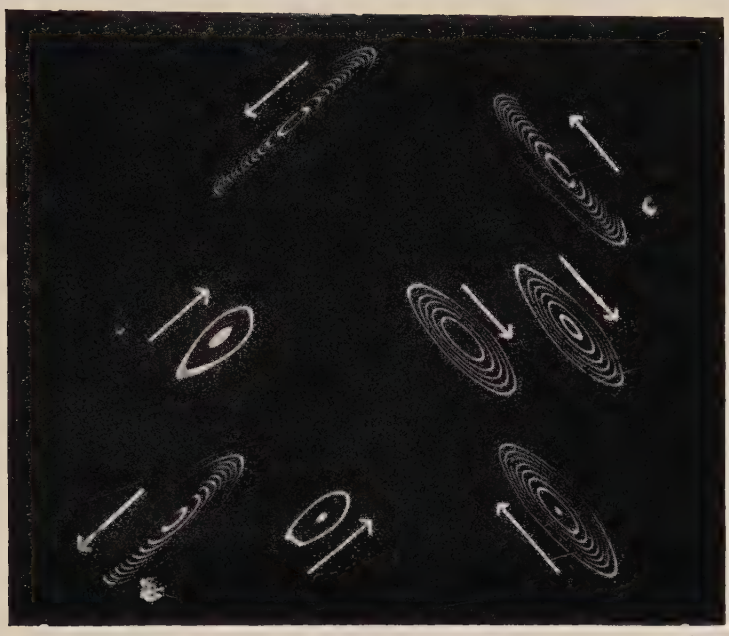


Fig. 1

×44

Fig. 2

×220

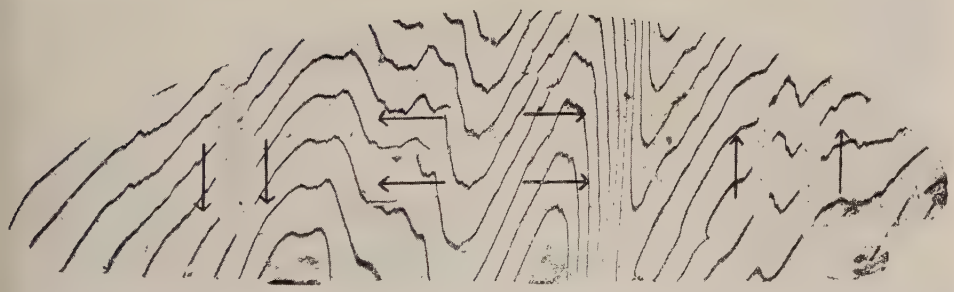
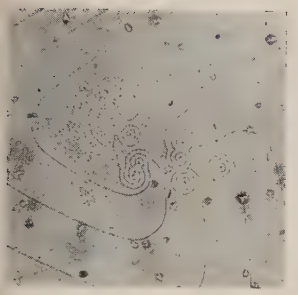
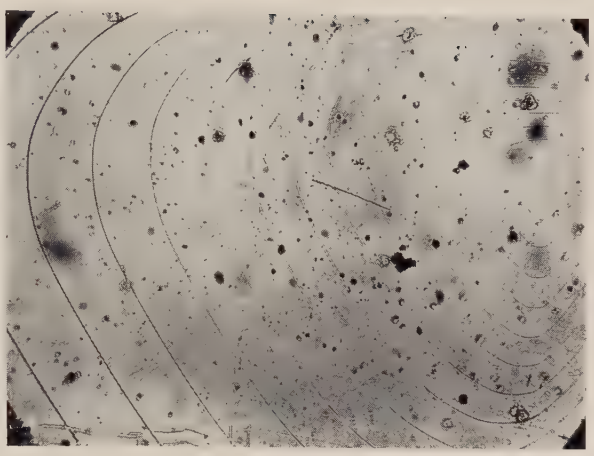


Fig. 3

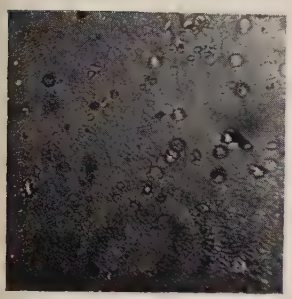
×36



(*b*) (×525)
SiC crystal



(*a*) (×75)
SiC crystal



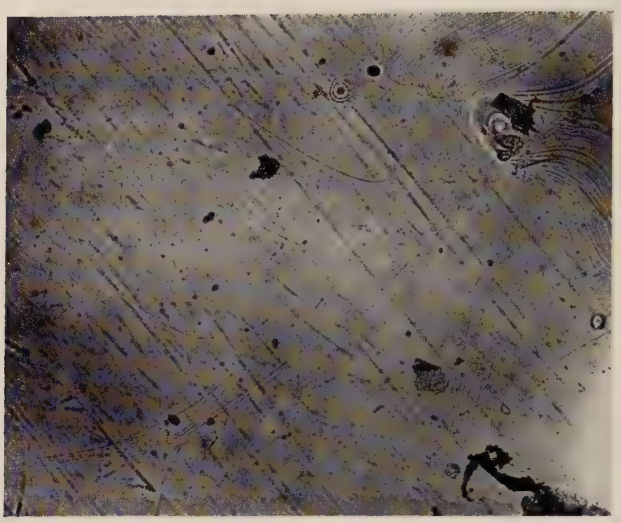
(*c*) (×525)
K_2CO_3—Na_2CO_3 (3 : 1)
5 min. etch

(*d*) (×525)
K_2CO_3—Na_2CO_3 (3 : 1)
10 min. etch

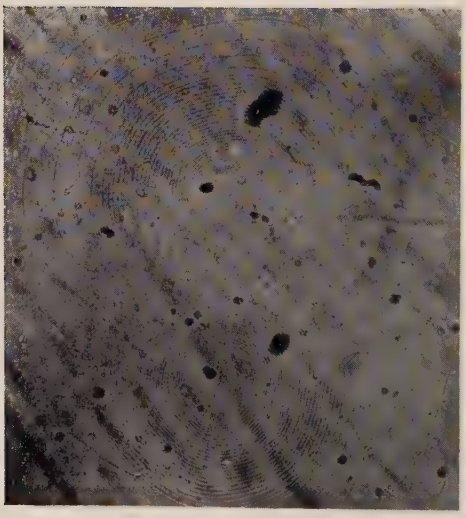
(*e*) (×525)
K_2CO_3—Na_2CO_3 (3 : 1)
40 min. etch

(*f*) (×525)
K_2CO_3—Na_2CO_3 (3 : 1)
80 min. etch

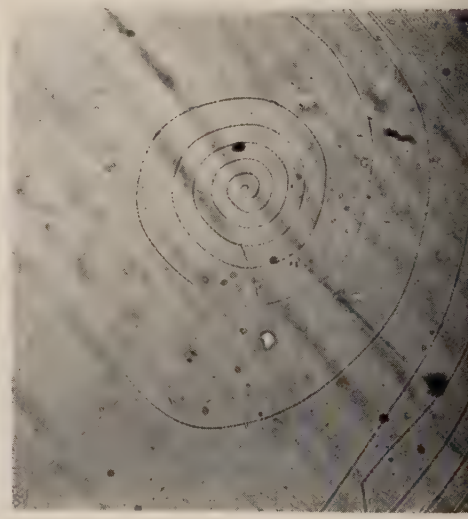
Simple Spirals



(*a*) (×75)
SiC crystal



(*b*) (×250)
Central area



(*c*) (×250
Upper right area

Complex Spirals

Phil. Mag. Ser. 7, Vol. 43, Pl. LXXV.

(*a*) Na_2CO_3 etch 10 min. (×75)

(*b*) Na_2CO_3 etch 20 min. (×75)

(*c*) Na_2CO_3 etch 40 min. (×75)

(*d*) Na_2CO_3 etch 80 min. (×75)

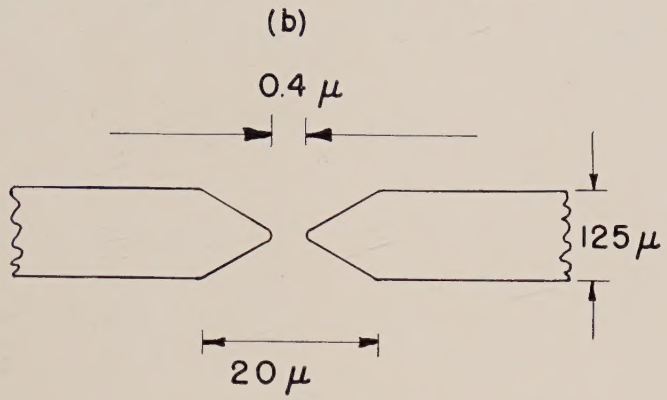
Etched Complex Spirals

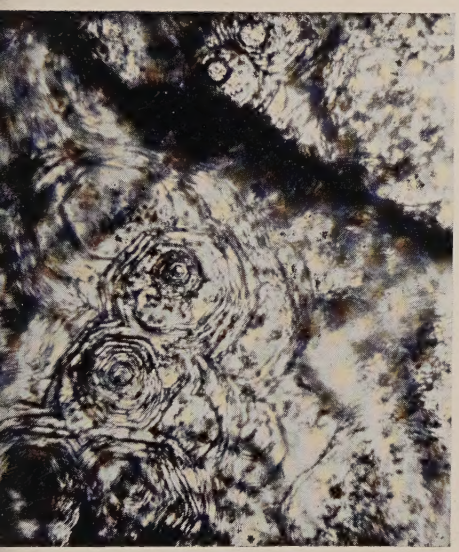
(*a*) Dark field (×25)

(*b*) Dark field (×25)

No visible Spirals

(*a*)
×788—diam. ∼ 0·4 μ





(*b*)
×525—diam. ∼ 2 μ

(*c*)
×125—diam. ∼ 240 μ

Holes in SiC